Are Ratings the Worst Form of Credit Assessment Except for All the Others?

ANDREAS BLÖCHLINGER and MARKUS LEIPPOLD*

ABSTRACT

We present a prediction model to forecast corporate defaults. In a theoretical model, under incomplete information in a market with publicly traded equity, we show that our approach must outperform ratings, Altman's Z-score, and Merton's distance to default. We reconcile the statistical and structural approaches under a common framework, i.e., our approach nests Altman's and Merton's approaches as special cases. Empirically, we cannot reject the superiority of our approach. Furthermore, the numbers of observed defaults align well with the estimated probabilities. Finally, with rank transforms, we obtain cycle-adjusted forecasts that still outperform ratings.

JEL classification: G01, G18, G24
Key Words: Calibration, Discrimination, Distance to Default, Z-score, Probit, Rank Transform

*Andreas Blöchlinger is with the Zürcher Kantonalbank and the University of Zurich. Markus Leippold is with the University of Zurich and the Swiss Finance Institute (SFI). We thank Karan Bhanot, Peter Christoffersen, Sergei Davydenko, Jin-Chuan Duan, Michael Gordy, Harald Hau, Liuren Wu, and the participants at the conference of the European Finance Association (EFA) 2012 in Copenhagen, the Seventh Annual Risk Management Conference in Singapore 2013, the Quant Europe 2014 Conference in London, the finance seminar at the University of Zurich, the Rotman School of Management, and the University of Toronto, for useful comments. We thank Basile Maire for augmenting and arranging the dataset used for our empirical analysis. The content of this paper reflects the personal view of the authors. In particular, it does not necessarily represent the opinion of the Zürcher Kantonalbank. Markus Leippold gratefully acknowledges financial support from the Swiss Finance Institute (SFI) and Bank Vontobel.

Not only has the role of rating agencies in financial markets been criticized, but also the regulators' policies concerning these agencies. In response to the recent financial crisis, politicians have aimd at weaning the financial industry from its dependence on ratings, and have therefore searched for more transparent alternatives. The Dodd–Frank act mandates that all financial agencies scan their regulations for references to credit ratings, and then remove and replace them with an appropriate alternative. But what alternatives do we have? We will give a theoretical answer for corporate ratings and support our theoretical arguments with an empirical application.

In the structural model of Merton (1974), Fischer, Heinkel, and Zechner (1989), or Leland (1994), the distance to default is a sufficient statistic for the prediction of a firm's default. The market-based distance to default is the number of standard deviations a firm is away from default and any additional data like ratings, financial ratios, asset prices, or macroeconomic variables cannot improve the forecasting power and is therefore marginally redundant. The empirical results of Duffie, Saita, and Wang (2007), Bharath and Shumway (2008), Campbell, Hilscher, and Szilagyi (2008), or Duffie, Eckner, Horel, and Saita (2009) indeed show the significant dependence of defaults on a firm's distance to default, but to a lesser extent also to further variables such as the firm's trailing stock return. However, using purely empirical arguments for favoring a particular forecasting model may not be satisfactory given the obvious features of default data like scarceness, dependencies, and sample selection biases. Duffie and Lando (2001) argue that in a market in which equity is not publicly traded, the distance to default cannot be accurately measured, and other covariates reveal marginal information about the firm's creditworthiness. For firms with publicly traded equity, however, Moody's KMV still entertains a commercial version of the Merton (1974) model in which the distance to default is the sole explanatory variable. There are no structural models available yet indicating why Moody's KMV-approach should not yield the most powerful default forecasts for firms with publicly traded equity, and therefore why it could not be regarded as an alternative to ratings.

In a world in which asset values are not perfectly observable, but equity may be traded in public capital markets, we demonstrate within a 1-period structural probit setup that an augmented model that combines Merton's market-based distance to default measure with the reduced-form statistical approach initiated by Altman (1968) must lead to default forecasts that are more powerful than the stand-alone approaches. Since rating agencies group borrowers into bins and since they neglect the information provided by the current state of the economy (a so-called through-the-cycle approach), their forecasts must be suboptimal as well. We will show, both theoretically and empirically, that over a fixed forecast horizon, the structural approach of Merton (1974), the reduced-form statistical approach of Altman (1968), Altman, Haldeman, and Narayanan (1977) as well as the expert approach applied by rating agencies lead to suboptimal default predictions in terms of discriminatory power.[1]

Discrimination, however, is just a relative measure to assess the performance of default forecasts. Calibration, on the other hand, addresses the absolute dimension of the default risk. Rating agencies explicitly state that they do not provide absolute measures of default risk. Instead, they provide a relative ranking of default risks from low to high.[2] Ranked data or ordinal measurements describe the order of the data. By definition, intervals and ratios between different ranks have no meaning. Consequently, ratings cannot provide answers to questions such as: How likely is an A rated borrower to default? How much more likely does a BB+ rated issuer default than a BBB- rated borrower does? For managing credit risk, such questions are naturally of pivotal importance.

To answer these questions, we need absolute measures in the form of estimated default probabilities. For instance, to map a distance to default into a probability of default (PD), we need the explicit distribution of asset returns. If we assume Gaussian distributed asset returns but true asset returns are fat-tailed, e.g., in the form of a Student's t-distribution, then we underestimate the true PDs of low risk firms. In other fields—such as weather forecasting or medical prognosis—the statistical comparison of estimated probabilities (e.g., precipitation probabilities, medical recovery probabilities) with realized frequencies is very well established (see e.g., Hosmer, Hosmer, le Cessie, and Lemeshow (1997)).[3] We can browse through academic papers on default prediction such as Shumway (2001), Chava and Jarrow (2004), Duffie, Saita, and Wang (2007), Bharath and Shumway (2008), Campbell, Hilscher, and Szilagyi (2008), Hilscher and Wilson (2012), or Duan, Sun, and Wang (2012), and see that goodness-of-fit is not addressed. In particular, if we make strong distributional or parametric assumptions, we may end up with a mathematically neat, elegant model, but with a lack of fit especially for the lowest or highest risk quintals, yet without adversely affecting its discriminatory power. That is, a model might perform well relatively but not absolutely.

The latest generation of default prediction models, as initiated by Duffie, Saita, and Wang (2007), is based on duration analysis, in which one models Poisson default intensities with stochastic covariates, and then one integrates over various forecast horizons to obtain the term structures of the default probabilities. However, to arrive at such term structures, one needs to impose quite strong assumptions on the time-series process of the covariates.[4] We will show that distributional assumptions that are too restrictive can lead to a lack of fit or a miscalibration. Instead, by fixing a forecast horizon and then resorting to a probit analysis, we will gain the flexibility to apply non-parametric statistical elements to arrive at powerful and calibrated forecasts without the need to specify the process of the covariates. We will demonstrate in particular that the non-parametric rank transformation is a powerful and useful statistical tool both to obtain calibrated default forecasts and to filter out the influence of the current economic state.

A calibrated model combines two important and testable properties. First, a calibrated model provides an unbiased estimate for the overall number of defaults. A portfolio of 1,000 borrowers with a mean default probability of 3% experiences on average 30 credit defaults, i.e., the PD interval between the default-risky portfolio and the default-free asset must be 3%. Second, calibrated forecasts differentiate correctly between low and high default probabilities. A sub-portfolio with a mean PD estimate of 4% has on average twice the default rate at the end of the forecasting period as a sub-portfolio with an average PD estimate of 2%, i.e., the PD ratio between the two sub-portfolios must be 2 to 1. Importantly, these two properties over different borrowers (cross section) must hold at each point in time or in every economic state, respectively (longitudinal section).

To the best of our knowledge, it has not been tested so far whether any default prediction model provides calibrated forecasts such that the interval lengths and the ratios between the PDs are correct at each point in a business cycle. Default dependencies inevitably bring the challenge of using the appropriate test statistics. This problem, which is further aggravated by the sparseness of the default data, led the Basel Committee on Banking Supervision (2005) to conclude that "statistical tests alone will be insufficient to adequately validate an internal rating system." This obstacle may be the explanation why most of the current literature does not fully analyze their models in terms of calibration. We will perform extensive calibration tests on a population of S&P rated borrowers with publicly traded equity shares with the help of newly developed test statistics by Blöchlinger (2012) and Blöchlinger and Leippold (2011).

Using financial statements and market information, we estimate our model with data from 1982 to 1999. Then, we test our model's out-of-sample forecasting performance in terms of calibration and discrimination for the period from 2000 to 2010. Our calibration tests show that our forecasts perform well absolutely in each year, i.e., at each point during the credit cycle, our forecasts are calibrated. Importantly, the calibration property does not come at the expense of discriminatory power. Thus, in terms of ranking the creditworthiness of the borrowers from best to worst, our model outperforms several benchmarks: Standard & Poor's corporate issuer ratings, the ζ-score of Altman, Haldeman, and Narayanan (1977), the Z-score of Altman (1968), and the distance to default model as discussed in Vassalou and Xing (2004). We provide a through-the-cycle version of our model to filter out macroeconomic effects and we base our default forecasts on this version only on relative rankings. Our cyclically adjusted version provides a new way to derive PDs since we resort to a comparative sample at each point in time. Our cycle-adjusted model is calibrated as well, and is at least as powerful as ratings.

Overall, we will show that we can make calibrated and powerful default forecasts that can be deemed as appropriate alternatives to corporate credit ratings. Unlike ratings, our forecasts fulfill some of the minimal requirements stated by the regulator in that our model is transparent, easily replicable, it only feeds on publicly available data, and it is no black-box. Last but not least, our adjusted model is by construction non-cyclical and since this model version is also calibrated, it can be used as input for the calculation of regulatory capital requirements.[5]

The rest of this paper is organized as follows: In Section I we discuss the concept of discrimination and calibration and how rank transformations can be useful to obtain both powerful, calibrated forecasts and to filter out macroeconomic influences. Section II derives our structural default model in which neither Merton's distance to default nor Altman's Z-score is sufficient, but must be combined.

Section III describes the data. Section IV presents our in-sample statistical estimation and Section V the out-of-sample statistical validation. Section VI concludes.

I. Powerful and Calibrated Forecasts with Rank Transforms

We assume that we are in a 1-period economy with the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ and we want to make a forecast for the binary default variable $Y \in \{0, 1\}$ whose realization is not known before the end of the period. The default forecast P to predict the future outcome Y is based on today's publicly available information set $\mathcal{G} \subseteq \mathcal{F}$, i.e., P is $\mathcal{G}$-measurable and Y is $\mathcal{F}$-measurable.

A. Discriminatory Power

To measure the discriminatory power of a predictor P, we resort to the Lorenz (1905) curve:

Definition 1 (Lorenz Curve): *The Lorenz curve of the predictor P is the two-dimensional graph*

$$\left(\mathbb{P}\left\{P \leq p\right\}, \mathbb{P}\left\{P \leq p \middle| Y = 1\right\}\right) \tag{1}$$

over all $p \in (-\infty, +\infty)$.

On the y-axis we have the fraction of defaulters among a given fraction of the population on the x-axis. If P and Y are stochastically independent, i.e., $\mathbb{P}\left\{P \leq p \middle| Y = 1\right\} = \mathbb{P}\left\{P \leq p\right\}$, then P is a naive predictor. The naive model has no discrimination ability at all, the Lorenz curve therefore corresponds to the 45 degree line, i.e., in this case we expect to observe x% of all defaulters among the x% best rated firms. The Lorenz curve of the predictor that knows ex-ante whether a firm fails, stays at zero up to the point where the x-coordinate equals the fraction of non-defaulting firms and then directly goes to the point $(1, 1)$. The closer a model's Lorenz curve is to this hypothetical curve, the better is its discriminatory power. The area above the Lorenz curve will serve as our summary statistic of the predictor's discrimination ability.

We assume that public information $\mathcal{G}$ is generated by the firm's asset value V, the nominal amount of debt F, and further variables as summarized in the vector X, i.e., $\mathcal{G} = \sigma(V, F, X)$. Inter alia, the vector X contains the full information needed to compute the Z-score variables of Altman (1968) and the ζ-score variables of Altman, Haldeman, and Narayanan (1977), but we assume that X and F do not contain the information needed to compute V. We will also work with the variable V^* which is the firm's asset value investors would be willing to pay at the beginning of the period if the firm's manager made public their private information. This is an extension to the case of incomplete information in the spirit of Duffie and Lando (2001). Under incomplete information, the variable V^* cannot be recovered by public information. Formally, we say V^* is not $\sigma(X, V, F)$-measurable.

In the structural model of Merton (1974), there is no marginal private information, i.e., $V^* = V$, and the firm defaults on its debt when the asset value at the end of the period, i.e., $V^* \exp(r)$, cannot cover the nominal amount of debt F, where r is the logarithmic asset return. Naturally, the asset return r is not known before the end of the period and is assumed to be Gaussian distributed with variance σ^2 and mean $\mu - 0.5\sigma^2$. Merton's distance to default measure DD is then by definition given by (see, e.g., Bharath and Shumway (2008), Eq. (6) with T=1 period):

$$DD = \frac{\log V - \log F + \mu - 0.5\sigma^2}{\sigma}, \tag{2}$$

and since r is Gaussian by assumption, the conditional default probability is

$$\mathbb{P}\{Y = 1 | X, V, F\} = \mathbb{P}\{V^* \exp(r) < F | X, V, F\} = \Phi(-DD) = \mathbb{P}\{Y = 1 | DD\}, \tag{3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standardized Gaussian variable. The first equality follows from applying the definition of default, since we have $Y = \mathbf{1}_{\{V^* \exp(r) < F\}}$, where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. The last equality expresses the sufficiency of DD, i.e., the conditional probability is only a function of DD. The marginal information provided by the full information set X, V, F is marginally irrelevant in making default forecasts: the single variable DD is sufficient.

In Merton's model, $\Phi(-DD)$ is the most powerful model among all predictors which are generated by public information. This result is a consequence of the sufficiency in (3) and the following lemma.

Lemma 1: *The conditional probability $\mathbb{P}\{Y = 1 | \mathcal{G}\}$ is the most powerful predictor among all predictors generated by the information set $\mathcal{G}$, i.e., the Lorenz curve of any other $\mathcal{G}$-measurable predictor must lie above the Lorenz curve of the conditional probability $\mathbb{P}\{Y = 1 | \mathcal{G}\}$.*

The proof of Lemma 1 and the proofs of all the following propositions can be found in Appendix A. Conditional probabilities based on smaller information sets, e.g., $\mathbb{P}\{Y=1|\,X\}$, or $\mathbb{P}\{Y=1\}$, must be less powerful than $\Phi(-DD)$. Hence, the Z-score of Altman (1968) or the ζ-score of Altman, Haldeman, and Narayanan (1977) must have lower discrimination ability if the assumptions of Merton (1974) hold true (since DD cannot be computed from X and F). Note, Y would be even more powerful than $\Phi(-DD)$, but Y is not known/measurable as of today. By definition of the Lorenz curve in (1), any strictly monotone transformation of a predictor must have the same discrimination ability as the original predictor, so that $\Phi(-DD)$ and $-DD$ produce exactly the same Lorenz curve.

B. Calibration

Calibration is a completely different concept from discrimination:

Definition 2 (Calibrated Forecasts): *The predictor P is calibrated if* $\mathbb{P}\{Y=1|\,P\}=P$.

Therefore, if the Gaussian assumption of Merton (1974) indeed holds true, the predictor $\Phi(-DD)$ is not only the must powerful predictor among all predictors that use publicly available data, but is calibrated as well. However, if asset returns are non-Gaussian, then $\Phi(-DD)$ is still the must powerful predictor but no longer calibrated. In particular, if r is fat-tailed, we may underestimate the true default probabilities for low risks if we nonetheless apply $\Phi(-DD)$. The naive predictor $\mathbb{P}\{Y=1\}$ on the other hand is calibrated but not as powerful as the miscalibrated predictor $\Phi(-DD)$. This example highlights the difference between calibration and discrimination. A predictor can be calibrated but be less powerful than a miscalibrated predictor and vice versa. Naturally, it is highly desirable to make forecasts with a predictor that is both calibrated and powerful.

We will show in the next section that $\Phi(-DD)$ is neither calibrated nor powerful even under Gaussian assumptions when asset values are not perfectly observable. That is, augmenting DD with accounting data such as those entering the ζ-score of Altman, Haldeman, and Narayanan (1977) improves the default forecast beyond the information provided by DD. In a non-Gaussian world we will even need additional transformations to obtain calibrated forecasts.

C. Rank Transformation

Rank transformations are used to transform data that do not meet the assumptions of normality. This non-parametric statistical procedure has been recommended as being robust to non-normal errors, resistant to outliers, and efficient for many distributions, as concluded by Conover and Iman (1981). Rank transformations have already proved to be useful in financial ratio analyses and logistic regressions, as reported by Kane and Meade (1998) and Kane, Richardson, and Meade (1998). In the context of default prediction, we will demonstrate that rank transformations offer an elegant way to perform calibrated predictions based only on relative rankings, which makes these forecasts directly comparable to ratings that are by definition only relative risk assessments. Overall, rank transformations allow us to filter out macroeconomic effects in a non-parametric way and we further expect the rank transformation to improve both the calibration and discrimination.

To rank-transform a particular variable X, we resort to a set of observations or a comparative sample, $\left\{X^{(1)},...,X^{(N)}\right\}$, respectively. We then define the rank transformation function of the variable X in analogy to an empirical cumulative distribution function (cdf):

$$R_X(x) = \frac{1}{N}\sum_{j=1}^{N}\mathbf{1}_{\left\{X^{(j)}\leq x\right\}}, \text{ for any } x \in (-\infty,\infty),$$

where we normalize by the size of the observation sample N so that $R_X(\cdot) : \mathbb{R} \to [0,1]$. The rank transform dampens dramatically the influence of outliers, since the rank transformation function is by definition bounded above by 1 and below by 0. We will use two different comparative samples to perform rank transformations. The first sample includes past observations over a whole business cycle (=panel data), the second sample includes only current observations (=cross-sectional data). To simplify notation, we denote the first transformation by $R_X(\cdot)$ and the second one by $r_{t,X}(\cdot)$. Hence, a value of $R_X(x) = 0.2$ means that we can expect to observe 20% of all observations of X over a business cycle to be less than or equal to x (=long term ranking). A value of $r_{t,X}(x) = 0.2$ means that at time t, 20% of the observations have the same or lower values than x (=current ranking at time t). In a downturn, the distance to default DD can be low relative to past observations (=long term ranking), but relatively high if we compare it just to the current set of issuers (=current ranking).

[Figure 1 about here]

Figure 1 illustrates both the long-term and the current rankings of DD for January 2009 and 2011. The current ranking or, equivalently, the empirical complementary cumulative distribution function (cdf) as of January 2011 dominates the complementary cdf as of January 2009. Hence, in the midst of the financial market crisis, the empirical p-quantile of DD was smaller than at the beginning of 2011 for any $p \in (0, 1)$. To give an example, as of January 2009, a borrower with a distance to default DD of 4.0 was clearly above the median value of 1.2 but neither over the long term with a median of 5.0 nor as of January 2011 with a median of 6.8.

[Figure 2 about here]

Figure 2 illustrates the rank and power transformations of four accounting variables used by Altman, Haldeman, and Narayanan (1977). In Panel B of Figure 2, we see the transformation for the ζ-variable "cumulative probability." If a corporation has no retained earnings, then the rank-transformed cumulative profitability variable has a value of 22.4%. Over a business cycle, 77.6% of firms are expected to have a retained earnings to total assets ratio of more than zero. Figure 2 illustrates the transformation of three further ζ-variables.

To make powerful and calibrated forecasts, we will apply rank and power transformations to obtain a credit score S and then use the Gaussian link function $\Phi(-S)$ to obtain an estimated default probability. The variable S can then be nicely interpreted as the number of standard deviations a borrower is away from default. Therefore, S is similar in spirit to the distance to default measure DD in Merton's model. The generic version of our augmented model takes the form

$$\begin{aligned}\mathrm{PD} &= \Phi(\kappa_0 + \kappa_1 R_{DD}(DD) + \kappa_2 R_{DD}(DD)^p + \kappa_3 \sum_{j=1}^{7} \eta_j R_j(\zeta_j)^q) \\ &= \Phi(-S),\ S = -\kappa_0 - \kappa_1 R_{DD}(DD) - \kappa_2 R_{DD}(DD)^p - \kappa_3 \textstyle\sum_{j=1}^{7} \eta_j R_j(\zeta_j)^q. \end{aligned} \tag{4}$$

The parameters p and q as well as κ_0, κ_1, κ_2, κ_3, and η_j, $j = 1, \ldots, 7$, are to be estimated from the data.[6] The variable $R_{DD}(DD)$ is the rank-transformed distance to default variable, and $R_j(\zeta_j)$ the rank-transformed accounting variable ζ_j.

Rating agencies claim that they take into consideration only relative risks. Therefore, the information set generated by ratings alone allows no inference about the state of the economy. As a direct consequence, if we are restricted to the information set provided by rating agencies, the conditional default probability for a given rating class must be the same in a downturn as well as in an upturn. A powerful model, however, must at least implicitly also consider the state of the economy.

To filter out the state of the economy from our general model in (4) and to make a fair comparison with agency ratings, we will perform additional forecasts based only on the relative rankings of S in (4). Thus, we consider only the current ranking $r_{t,S}$, and make no comparisons to past observations or different economic states. The current median value of S within the comparative sample can be higher or lower than last year's or even yesterday's median. We then make forecasts on this restricted information set as follows:

$$\begin{aligned} \mathrm{PD} &= \Phi(c_0 + c_1\, r_{t,S}(S) + c_2\, r_{t,S}(S)^v) \\ &= \Phi(-\bar{S}),\ \ \bar{S} = -c_0 - c_1\, r_{t,S}(S) - c_2\, r_{t,S}(S)^v. \end{aligned} \tag{5}$$

Note, we estimate here the conditional default probability $\mathbb{P}\{Y = 1|\,\mathcal{G}\}$ in Lemma 1 when the information set $\mathcal{G}$ consists only of relative rankings. The input is ordinal data, the outcome is an absolute measure. This novel approach to derive PDs needs the current, comparative sample at each point in time. Therefore, if a borrower is, say, in the 11th percentile, i.e., $r_{t,S}(S) = 0.11$ in an upturn as well as in a downturn, then the issuer will have the same conditional default probability in (5) in both economic states even though that the conditional default probability in (4) can be lower in an upturn as compared to a downturn. The same is true for a borrower in, say, the 56th percentile, so that the PD ratio between borrowers in the xth percentile and issuers in the yth percentile is constant over time for any $x, y \in (0, 1)$. Hence, both the PD levels and the PD ratios in (5) will remain unchanged over the business cycle. In the words of Blöchlinger and Leippold (2011), the PD function in (5), unlike the PD function in (4), has the same level calibration and the same shape calibration at each point in time. Again, $\bar{S}$ can be nicely interpreted as the number of standard deviations a borrower is away from default when we abstract from the current state of the economy.

The time series of many financial figures have no unit root, and in particular, the distance to default measure DD can be assumed to be mean reverting. This observation was characterized by Duffie, Saita, and Wang (2007) as "leverage targeting, by which corporations pay out dividends and other forms of distributions when they achieve a sufficiently low degree of leverage, and conversely attempt to raise capital and retain earnings to a higher degree when their leverage introduces financial distress or business inflexibility." In a good state of the economy, the median borrower's DD is higher than the long-term median, the median default probability is also lower, and firms can make larger dividend payments to their shareholders, which causes reductions of the distances to default and thereby increases the default probabilities again. The current ranking $r_{t,DD}$, unlike the long-term ranking R_{DD}, filters out these swings over the business cycle, so that the median of $r_{t,DD}$ is 0.5 at each point of time t. That is why we call the conditional PD in (4) our point-in-time, or, more simply, the **PIT** model, and the conditional PD based on relative rankings in (5) will be called the through-the-cycle, or, for short, the **TTC** model.

By Lemma 1, the discriminatory power of our **TTC** predictor on the restricted information set in (5) must be less powerful than the **PIT** predictor that can resort to the larger information set in (4). It is a well-known statistical property that the conditional expectation on a information set $\mathcal{G}$ has greater variance than the conditional mean on a subset of $\mathcal{G}$. The forecasts of model **PIT** are therefore expected to be more powerful but more volatile or less stable than the predictions from **TTC**. We will show in the following empirical analysis that even when we are limited to the information set of the current rankings of S, our **TTC** forecasts are still at least as powerful as ratings. But first we show, in a structural model, that augmenting the distance to default measure with additional data improves the prediction accuracy.

II. Reconciling Merton's and Altman's Approaches

Augmenting Merton's distance to default model with additional variables may seem an ad-hoc fix for incomplete information. Empirically, Bharath and Shumway (2008) already show that the Merton model does not produce a sufficient statistic for the probability of default. However, we can put the augmentation on a sound theoretical basis.

Since we are interested in making default predictions over a fixed forecast horizon, we consider a 1-period structural default model. We borrow the idea of Duffie and Lando (2001) that a firm's asset value is only observable with noise. Duffie and Lando (2001) derive the term structure of credit spreads when equity is not traded in public markets. In our model, both equity and debt may or may not be traded in public financial markets. Thus, our population explicitly includes firms with publicly traded equity. Our focus is on the probability that a firm is incapable of repaying its debt at the end of the forecast horizon.

In our structural model, we assume that the variable V^* corresponds to the price that market participants would be willing to pay today for the firm's assets when the firm's managers publish their private information. Both debt and equity holders have no access to managers' private information and must estimate today's asset value V^* based on publicly available information. Managers are precluded by insider-trading regulation from trading in public capital markets. Managers cannot be selective as to how they release information. This means that equity holders cannot have more information than debt holders. All market participants receive managerial information at the same time. Managers will initiate the liquidation of the firm if the end of period asset value, $V^* \exp(r)$, fails to meet the debt payment F, i.e., $Y = \mathbf{1}_{\{V^* \exp(r) < F\}}$ denotes the default indicator, r is the logarithmic asset return.

Assumption 1: *The market participants' estimate of today's asset value V^* using only public information is*

$$\log V = \log V^* + U,$$

where V is the market's estimate and U is the estimation error. The estimate V for the asset value V^* is unbiased.

We further assume that the mean asset return and the mean asset value are a function of the vector X, which may contain figures such as the risk-free rate, book-to-market ratio, turnover, market capitalization, book value of equity, current ratio, industry indicators, state of the economy, GDP, or Fama–French factors. For instance, the mean asset return can be the risk free return r_f plus a premium for factor risk. Thus, we make the following distributional assumptions:

Assumption 2: *The logarithmic nominal amount of debt $\log F$, the vector X, as well as $\log V^*$, r, and U follow a multivariate Gaussian distribution. The three variables V^*, r, and U, given the variables*

X and F, are supposed to be mutually independent:

$$\begin{pmatrix} \log V^* \\ r \\ U \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma^\top X \\ \lambda^\top X - \sigma^2/2 \\ -\alpha^2/2 \end{pmatrix}, \begin{pmatrix} \delta^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \alpha^2 \end{pmatrix}\right), \tag{6}$$

whereas $\delta, \sigma, \alpha > 0$.

Given joint normality, the independence assumption in (6) is without loss of generality. Under the above assumptions, we can derive the following proposition:

Proposition 1: *Given Assumptions 1 and 2 and conditional on the public information generated by F, V, and X, the default probability $\mathbb{P}\{Y = 1 | F, V, X\}$ is given by*

$$PD = \Phi\left(\beta_0 + \beta_1 \log F + \beta_2^\top X + \beta_{DD}\, DD\right) = \Phi\left(Z + \beta_{DD}\, DD\right), \; DD = \frac{\log\left(\frac{V}{F}\right) + \mu - \frac{1}{2}\sigma^2}{\sigma}, \tag{7}$$

where $\mu = \lambda^\top X$ as well as

$$\beta_1 = \frac{1}{\sqrt{\sigma^2 + \delta^2 - \frac{\delta^4}{\delta^2 + \alpha^2}}} \frac{\alpha^2}{\delta^2 + \alpha^2}, \; \beta_0 = \frac{1}{2}(\sigma^2 - \delta^2)\beta_1, \; \beta_2 = -\beta_1\left(\lambda + \gamma\right), \; \beta_{DD} = -\sigma\frac{\delta^2}{\alpha^2}\beta_1,$$

and $Z := \beta_0 + \beta_1 \log F + \beta_2^\top X$. Conditional on public and private information, the default probability $\mathbb{P}\{Y = 1 | F, V, X, V^*\}$ is given by

$$PD^* = \Phi\left(-DD^*\right), \; with\; DD^* = \frac{\log\left(\frac{V^*}{F}\right) + \mu - \frac{1}{2}\sigma^2}{\sigma}. \tag{8}$$

Hence, under the inclusion of private information, the distance to default measure DD^* is again sufficient. However, given only public information, the distance to default DD is no longer sufficient even though we have used an unbiased estimate for V^* in the form of V to compute DD. That is, the default probability in (7) is represented as a function of both Merton's distance to default measure DD and a score component Z, similar in spirit to Altman's Z-score. We can show that only under specific assumptions is one of these two approaches sufficient to build a default forecast. We first consider the case when the asset value V^* can be perfectly observed, i.e., when we let $\alpha \to 0$. Then we have $\beta_0 = \beta_1 = \beta_2 = 0$, $\beta_{DD} = -1$, and the PD reduces to Merton's original PD:

$$\text{PD} = \Phi\left(-DD\right). \tag{9}$$

Hence, the Z-score component becomes totally irrelevant in predicting defaults in the case of complete information. Second, if the estimated asset value V is extremely noisy, $\alpha \to \infty$, then $\beta_{DD} \to 0$ and we have a PD that is only a function of the Z-score:

$$\mathrm{PD} = \Phi\left(\beta_0 + \beta_1 \log F + \beta_2^\top X\right) = \Phi\left(Z\right), \tag{10}$$

with $\beta_1 = \left(\sigma^2 + \delta^2\right)^{-0.5}$ and β_0, β_2 as given in Proposition 1. Importantly, the distance to default measure drops out in this case.

Given Assumption 2, the variables $\log V$, $\log F$, and X are multivariate Gaussian. Hence, Z and DD follow a bivariate Gaussian distribution due to the linear transformation,

$$Z \sim N(\mu_Z, \sigma_Z^2),\ DD \sim N(\mu_{DD}, \sigma_{DD}^2),$$

with correlation coefficient ρ. If the asset values can be perfectly observed ($\alpha \to 0$), then Z is degenerate and equal to 0 with $\mu_Z = 0$ and $\sigma_Z = 0$. In Proposition 1, we have the default probability conditional on the situation when both Z and DD are known. In the proposition below, we derive the default probabilities for the case when both Z and DD are unknown or when only one of these variables is known.

Proposition 2: *Given Assumptions 1 and 2 and conditional on knowing only the measure DD, Merton's adjusted probability of default is given by*

$$\mathbb{P}\left\{Y = 1 |\, DD\right\} = \Phi\left(\frac{\mu_Z - \rho\frac{\sigma_Z}{\sigma_{DD}}\mu_{DD} + DD\left(\beta_{DD} + \rho\frac{\sigma_Z}{\sigma_{DD}}\right)}{\sqrt{1 + \sigma_Z^2\left(1 - \rho^2\right)}}\right) = \Phi\left(b_0 + b_{DD} DD\right), \tag{11}$$

where

$$b_0 = \frac{\mu_Z - \rho\frac{\sigma_Z}{\sigma_{DD}}\mu_{DD}}{\sqrt{1 + \sigma_Z^2\left(1 - \rho^2\right)}},\ \textit{and}\ b_{DD} = \frac{\beta_{DD} + \rho\frac{\sigma_Z}{\sigma_{DD}}}{\sqrt{1 + \sigma_Z^2\left(1 - \rho^2\right)}}.$$

Similarly, conditional on knowing only Z, Altman's probability of default is given by

$$\mathbb{P}\left\{Y = 1 |\, Z\right\} = \Phi\left(\frac{\beta_{DD}\mu_{DD} - \rho\beta_{DD}\frac{\sigma_{DD}}{\sigma_Z}\mu_Z + Z\left(1 + \rho\beta_{DD}\frac{\sigma_{DD}}{\sigma_Z}\right)}{\sqrt{1 + \beta_{DD}^2\sigma_{DD}^2(1 - \rho^2)}}\right). \tag{12}$$

If no source of information at all is available, we can only apply the naive probability of default:

$$\mathbb{P}\{Y=1\} = \Phi\left(\frac{\mu_Z + \beta_{DD}\,\mu_{DD}}{\sqrt{1+\sigma_Z^2+\beta_{DD}^2\sigma_{DD}^2+2\rho\beta_{DD}\sigma_{DD}\sigma_Z}}\right), \tag{13}$$

which cannot discriminate between low and high risks.

By construction, the predictor $\Phi(-DD^*)$ based on public and private information in (8), our augmented PD, $\Phi(Z+\beta_{DD}\,DD)$, based solely on public information in (7), Merton's adjusted PD in (11), Altman's PD in (12), as well as the naive PD in (13), are all calibrated, since they satisfy Definition 2. The Lorenz curves for these five predictors as defined in (1), however, will in general look different. According to Lemma 1, $\Phi(Z+\beta_{DD}\,DD)$ in (7) is more powerful than Merton's PD and Altman's PD, as well as the naive PD. On the other hand, $\Phi(-DD^*)$ in (8) can discriminate even better than $\Phi(Z+\beta_{DD}\,DD)$.

These differing discrimination abilities can be explained intuitively as follows: Merton's PD neglects the information provided by Z. If two observations have the same distance to default measure DD, Merton ranks them as equally likely to default even though they have different default probabilities due to different values of Z. Hence, among the group of observations with the same distance to default DD, we could still make a refinement of the ranking by exploiting the information in Z. Hence, our augmented PD in (7) necessarily outperforms Merton's PD in (11) in terms of discriminatory power.

Merton's adjusted PD, $\Phi(b_0+b_{DD}DD)$ in (11), is a monotone transformation of Merton's original PD, $\Phi(-DD)$ in (9). Hence, both predictors provide the same relative ranking. In terms of discriminatory power both work equally well and produce exactly the same Lorenz curve. However, if asset values are not perfectly observable, then $\Phi(-DD)$ unlike $\Phi(b_0+b_{DD}DD)$ is not calibrated and some risks are underestimated while other risks are overestimated. Indeed, a clear pattern emerges.

Proposition 3: *If asset values are not perfectly observable, i.e., $\alpha>0$, then Merton's original PD, $\Phi(-DD)$, underestimates low default risks and overestimates high default risks.*

Proposition 3 states that if we wrongly assume asset values to be perfectly observable, we will necessarily observe more defaults than predicted for low-risk issuers (with a distance to default above a critical value) and the mean default rate will be lower than forecast for high-risk issuers. Hence, whenever asset values are not perfectly observable, we must first apply a linear transformation to DD to obtain calibrated forecasts.

Our default probability in (7) is a continuous function of Z and DD. Any grouping of forecasts into discrete classes, as is done by rating agencies such as S&P, must result in an inferior ranking. Issuers with different default probabilities end up in the same rating class. Therefore, if we group observations into discrete rating classes, we lose discriminatory power. We can no longer distinguish between low and high risks within a given rating class. This loss of power due to discretization seems to be quite acute between S&P's BBB- and BB+ as the observed 1-year default frequency for the former group is 0.06% and 0.36% for the latter, as plotted in Panel B of Figure 3. A PD ratio of six to one for two neighboring rating notches is quite sizeable.

Furthermore, rating agencies admittedly disregard the state of the economy. They neglect part of the information provided by X. Therefore, ratings can no longer be as powerful as the augmented predictor in (7), unless they can incorporate private information to make up for this shortcoming. However, insider trading regulation implies that rating analysts cannot be privy to the managers' private information ahead of the general public.

From our Gaussian assumptions, the end of period log asset value, $\log V^* + r$, given public information, is again Gaussian. As a consequence, the optimal predictor in the form of the conditional default probability in (7) is the mapping of a linear function of Z and DD via the Gaussian link function $\Phi(\cdot)$. Such a specification corresponds to a probit setup and the model parameters can be estimated by maximizing the likelihood function. The best performing model can be selected by likelihood ratio tests. However, under non-normality, the functional form of the conditional probability will look different. We will see in the following empirical analysis that we can construct better fitting models by foregoing the distributional restrictions imposed by Assumption 2.

III. The Dataset

Our population includes all non-financial corporations rated by S&P between 1982 and October 2011 that have publicly traded equity shares.[7] We are not aware of any reason why our findings for S&P should not equally apply to the populations rated by Moody's or Fitch. Financial ratios, equity, and debt data was retrieved from Bloomberg, and we divided our data into an estimation sample and a validation sample. The estimation sample ranges from 1982 to 1999 and serves to select the relevant predictor variables and to estimate the parameters of the forecasting models. The validation sample ranges from 2000 to 2011 and allows us to test the forecasting performance of different models and to benchmark them against S&P ratings. Our objective is to forecast whether a firm defaults within one year. Our S&P data ends in October 2011, hence the last date for which we know whether a given firm has defaulted within one year is October 2010.

To avoid overlapping time windows, we select the data at one particular month of a year, i.e., we include S&P ratings, accounting data, and distance to default measures by the end of October for every firm and year. Publicly held corporations in the US are required to report earnings to the SEC within 45 days of the end of their first three fiscal quarters. However, most firms release their results much earlier through press releases. Thus, the end of October is a time when the so-called earnings season is virtually over and the bulk of companies has released their third quarter earnings reports.[8]

At the end of each October and for each firm, we observe the predictor variables and whether S&P downgraded the firm to D ("in default") one year later. Since the distance to default is market-based, it changes every day. In contrast, most of the variables of the Z-score and ζ-score are based on balance sheet data, which are available, at most, quarterly. For balance sheet data, we use the latest available report. In total, the estimation sample contains 7,684 firm-years, and the validation sample contains 16,727 firm-years. We summarize the dataset used for our empirical study in Table I.

[TABLE I about here]

Moody's KMV uses a proprietary version of the Merton (1974) distance to default model. The reconstruction of the distance to default measure with publicly available data can be found in Appendix B. Fink (2003) presents a table with Moody's KMV estimated PDs and asset volatilities for 95 companies as of August 2000. As a sanity check, Bharath and Shumway (2008) compare their estimated volatilities and PD estimates of 80 companies with these numbers. The rank correlation of Spearman (1904) between their estimates and Moody's KMV is 79% for PDs and 57% for asset volatilities. The same sanity check is performed with our dataset. In total, there are 78 firms for which Spearman's rank correlation between our estimates and Moody's KMV can be calculated. For PDs, the rank correlation is 83.4%, and 90.3% for asset volatility estimates. Hence, we can safely conclude that the distance to default model here is similar to the version of Bharath and Shumway (2008), demonstrating that we are able to capture much of the information in Moody's KMV's estimates.

IV. In-Sample Estimation

For our empirical analysis, we estimate the models introduced above under different specifications by maximizing the likelihood function. Then we look at the time series of eight selected companies, of which one-half have defaulted, to gain a qualitative impression of our PD estimates.

A. Model Estimation and Selection

First, we use Merton's original PD in (9):

$$\mathbf{M1}: \qquad \mathrm{PD} = \Phi\left(-DD\right).$$

Since **M1** is a purely structural specification, no parameters need be estimated from the data. However, we can determine the likelihood value of this specification for comparative statistics as reported in Table II. We then perform a probit analysis to estimate the parameters b_0 and b_{DD} for Merton's adjusted PD in (11). We obtain the following PD estimate for this specification:

$$\mathbf{M2}: \qquad \mathrm{PD} = \Phi\left(b_0 + b_{DD}DD\right) = \Phi\left(-1.57 - 0.34DD\right).$$

In Table II, the large t-values indicate that the constant b_0 becomes significantly different from zero and that the sensitivity parameter b_{DD} is significantly different from minus one. By comparing the likelihoods, we find that model **M2** has a significantly better fit than **M1**. This finding is an empirical confirmation of Proposition 3, that Merton's original PD systematically underestimates low risks and overestimates high risks.

[TABLE II about here]

Since it is well-documented that asset returns are fat-tailed, we expect that rank and power transformations applied to the distance to default variable DD will improve the goodness of fit. Hence, we estimate the following model:

$$\begin{aligned}\mathbf{M3}: \qquad \mathrm{PD} &= \Phi\left(a_0 + a_1 R_{DD}(DD) + a_2 R_{DD}(DD)^p\right) \\ &= \Phi(-D),\ D = 4.49 - 2.64\, R_{DD}(DD) - 1.63\, R_{DD}(DD)^{30},\end{aligned}$$

where a_1, a_2, a_3, and p are parameters estimated from the data, and $R_{DD}(DD)$ denotes the rank-transformed distance to default variable. As shown in Table II, the transformation of $R_{DD}(DD)$ to the power of 30 has a higher likelihood than the transformations to the power of 10 (model **M4**) and 50 (model **M5**). Importantly, **M1** to **M5** all produce the same Lorenz (1905) curve, as they are simply a monotone transformation of the distance to default measure into a 1-year default probability. All perform equally well in terms of discrimination. However, in terms of calibration or goodness of fit, they differ hugely. Panel A of Figure 3 shows the different mappings of DD into an estimated default probability for **M1**, **M2**, **M3**. The log likelihood values of these three models differ significantly with values of -1158.0, -330.3, and -321.3.

In other words, the structural and distributional assumptions of model **M1**, and **M2** as stated in Section II are clearly rejected. On the one hand, Merton's complete information assumption in **M1** is rejected in favor of incomplete information as reflected by **M2**. On the other hand, Assumption 2 under multivariate Gaussian variables is definitely too restrictive and does not provide well fitting models or calibrated forecasts, since **M3** statistically outperforms **M2**. In summary, **M1** and **M2** suffer from a lack of fit, both models are therefore miscalibrated and rejected in favor of **M3**. Panel B of Figure 3 shows the mapping of S&P rating classes into default probabilities estimated by simple count statistics, i.e., the number of defaulters at the end of the observation period divided by the number of observations at the beginning for each rating class

[Figure 3 about here]

Table III shows the results from the estimation of probit models for the Z-score and the ζ-score variables. Since these models are not nested, we cannot resort to the likelihood ratio to select the best performing model. Instead, we base our selection on the information criterion of Akaike (1974). The model with the ζ-score variables statistically outperforms the model with the Z-score data. To limit the influence of outliers, we apply again the rank transformation to the original financial ratios. With our rank-transformed variables, we obtain higher likelihoods than by simply Winsorizing the variables, as previously suggested by Campbell, Hilscher, and Szilagyi (2008). Some ζ-variables are statistically insignificant. Therefore, we exclude these variables from the subsequent analysis and we only use debt service (ζ_3), cumulative profitability (ζ_4), capitalization (ζ_6), and company size relative to median value (ζ_7). Besides the rank transformation, an additional square root transformation provides an even better fit of the data. Note that instead of using the original size variable in dollars, as suggested by Altman, Haldeman, and Narayanan (1977), we divide it by the current median asset value of the underlying population of borrowers. Otherwise, the size variable would have a unit.

[TABLE III about here]

Among all accounting-based prediction models in Table III, we select the specification given by

$$\begin{aligned}\mathbf{A1}: \quad \mathrm{PD} &= \Phi(\eta_0 + \sum_{j=1}^{7} \eta_j R_j(\zeta_j)^q) \\ &= \Phi\left(-Z\right), \; Z = -0.22 + 0.99\sqrt{R_3(\zeta_3)} + 0.67\sqrt{R_4(\zeta_4)} + 1.88\sqrt{R_6(\zeta_6)} + 1.31\sqrt{R_7(\zeta_7)},\end{aligned}$$

where Z is a credit score in analogy to Altman, Haldeman, and Narayanan (1977), and $R_3(\zeta_3)$, $R_4(\zeta_4)$, $R_6(\zeta_6)$, and $R_7(\zeta_7)$ are the rank-transformed variables, Z describes the number of standard deviations a firm is away from default when only the variables of Altman, Haldeman, and Narayanan (1977) are used as the information set. Since the financial ratios are definitely non-Gaussian, it was to be expected that models with rank-transformed financial ratios would outperform those where the ratios are simply Winsorized. Therefore, **A1** represents the non-Gaussian empirical equivalent of the conditional probability derived in (12).

As shown in Table II, Merton's distance to default model, **M3**, has a significantly higher likelihood value than Altman's ζ-score model, **A1**, which has two more parameters, -321.3 vs. -351.2. Therefore **M3** is statistically superior to **A1**. In fact, by examining the likelihoods and the number of parameters, we find that all distance to default models in Table II, except Merton's original model **M1**, are statistically superior to **A1**.

[Figure 4 about here]

From Figure 4, we observe that the ζ-score and the distance to default are positively correlated, but not perfectly so. Spearman's rank correlation shows a coefficient of 64.1% in the estimation sample (Panel A) and 57.1% in the validation sample (Panel B). As expected, the majority of defaulters are concentrated in the tenth deciles, both in the estimation sample and the validation sample. Hence, as the correlation is not perfect, there is hope for improvement when both variables are combined into an augmented default risk measure. Therefore, we implement different combinations of market-based and accounting variables and search for the best model specified by Equation (4). We call it the point-in-time (**PIT**) model:[9]

$$\begin{aligned}\textbf{PIT}: \qquad \mathrm{PD} &= \Phi(\kappa_0 + \kappa_1 R_{DD}(DD) + \kappa_2 R_{DD}(DD)^p + \kappa_3 \sum_{j=1}^{7} \eta_j R_j(\zeta_j)^q) \\ &= \Phi\left(-S\right),\ S = 2.88 - 1.79 R_{DD}(DD) - 1.27 R_{DD}(DD)^{30} + 0.40\, Z. \qquad (14)\end{aligned}$$

Again, by maximum likelihood we obtain parameter estimates for κ_0, κ_1, κ_2, κ_3, and p. The variable S is a standardized credit score. Note, for the aggregated variable Z, we do not re-estimate η_j or q, but we take the values from model **A1** as given. We justify this procedure by considerations of statistical robustness and parsimony. As the t-value for the ζ-score in Table III suggests, augmenting the model with Z significantly increases its information content. The likelihood ratios demonstrate that the model **PIT** (log likelihood of -311.7) is statistically superior to the best performing distance to default model **M3** (log likelihood of -321.3) and the best performing ζ-score model **A1** (log likelihood of -351.2). This empirical result was expected from Proposition 1, since asset values cannot be observed but must be estimated, with the consequence that the distance to default cannot be a sufficient statistic.

In addition, we also implement our business-cycle adjusted default probability model, which we call the through-the-cycle (**TTC**) model:

$$\begin{aligned}\textbf{TTC}: \qquad \mathrm{PD} &= \Phi(c_0 + c_1\, r_{t,S}(S) + c_2\, r_{t,S}(S)^v) \\ &= \Phi\left(-\bar{S}\right), \ \bar{S} = 3.37 - 0.83\, r_{t,S}(S) - 1.99\, r_{t,S}(S)^{10}. \end{aligned} \tag{15}$$

where $r_{t,S}(S) \in [0,1]$ is the current ranking at time t of our point-in-time distance S in Equation (14). By construction, the mean PD of forecasting model **TTC** is constant over time and since **TTC** is based on a smaller information set, its optimized likelihood value of -332.5 is significantly lower than **PIT**. Thus, due to the larger information set, the mean default probability of model **PIT** tends to be higher before an expected downturn and lower before an anticipated expansion. Under the **TTC** model, the population has the same mean default probability of 1.34%, the same median PD of 0.16%, and the same minimum (maximum) PD of 0.038% (29.1%) at each point in time. Hence, the default probability in Equation (15) corrects the default probability obtained in Equation (14) for business-cycle effects. A conditional PD based on current rankings only allows no inference on the state of the economy. By filtering out business cycle effects, we obtain conditional default probabilities that look "through the cycle."

Merton's distance to default DD is actually meant to express the number of standard deviations a firm is away from default. However, we clearly reject this calibration hypothesis since D from model **M3** is the more appropriate distance measure in terms of standard deviations away from default. The distance measures Z from model **A1** is supposed to be calibrated as well, but judging from Akaike's information criterion, it seems to be significantly less powerful than D. Furthermore, based on our estimation results, the **PIT** score S,

$$S = 2.88 - 1.79\, R_{DD}(DD) - 1.27\, R_{DD}(DD)^{30} + 0.40\, Z, \tag{16}$$

is calibrated and the most powerful distance measure. As a consequence, the **TTC** score $\bar{S}$,

$$\bar{S} = 3.37 - 0.83\, r_{t,S}(S) - 1.99\, r_{t,S}(S)^{10},$$

expresses the number of standard deviations away from default in terms of the relative risk corrected for the state of the economy. Note that, even though our score $\bar{S}$ is based solely on relative risk assessments, we obtain nonetheless a conditional default probability estimate in the form of $\Phi(-\bar{S})$. Hence, in contrast to the forecasts provided by rating agencies, we obtain calibrated forecasts.

There are different views on whether credit ratings should be point-in-time or through-the-cycle forecasts. Rating agencies filter out the higher bankruptcy risk during recessions, that is, agency ratings are through-the-cycle ratings. However, market participants arguably want to anticipate the increased default risk even if the heightened default intensity is expected to be only transitory and not permanent. This view is also implicitly supported by the Dodd–Frank Act.[10] On the other hand, there are valid arguments for non-cyclical capital requirements, which must be based on cycle-adjusted default probabilities. In any case, we now have a point-of-time version, **PIT**, and a through-the-cycle version, **TTC**, of our model.

We remark that we floor the estimated default probabilities at 0.02% in **M3**, and **A1**, to further reduce the effects of outliers. We also floor and cap the PDs of model **PIT** by maximizing the likelihood function once more to obtain a floor of 0.031% and a cap of 37% under an optimized likelihood value of -310.1 (by keeping the other parameters fixed). That is, the in-sample likelihood values improve only insignificantly by flooring and capping (remember, without cap and floor we have a likelihood value of -311.7), but we think that very low and very high probability estimates are more likely to be driven by noise than by signal, so we use the floored and capped PD estimates in the following. Note that only a small fraction of less than 1% of the estimation sample is affected by the cap.

S&P ratings are by definition ordinal measures. We employ simple count statistics to map monotonically the alphabetic rating scale onto a scale between zero and one to make them comparable to PD estimates as shown in Panel B of Figure 3. In particular, for each rating class, we sum the number of observations over all months in the estimation sample in the denominator, and we sum the corresponding number of defaulters one year later in the numerator. This quotient is an estimate for a "through the cycle default probability" for the corresponding rating class. Given only ratings data, we have no information on the state of the economy. As a consequence, a borrower with the same rating in two different states of the economy must have the same conditional PD in both states unless we were to supplement the ratings with marginal information that is not provided by the rating agencies. To have strictly positive values for each rating class and to be strictly monotone across classes, we assign an estimate by interpolation/extrapolation to the lowest risk classes, whereas the AAA class is mapped onto a value of 0.001%. This monotone PD mapping allows us to plot time series of S&P ratings, **M3**, **A1**, **PIT**, and **TTC**, on a common scale for selected companies.

B. Probability Estimates for Selected Firms

Figure 5 shows the time series of 1-year ahead forecasts for four defaulting firms: Enron, Parmalat, Delphi, and General Motors. The performance of S&P is mixed at best. They were usually slow in downgrading a subsequent defaulter. Our observation is consistent with several surveys which show that market participants believe that rating agencies are slow in responding to changes in credit quality.[11] The agencies usually reply to this critique that their ratings are "through the cycle," with the intention of measuring credit quality over long investment horizons. Rating stability, however, is often in conflict with rating timeliness.

[Figure 5 about here]

From Figure 5 we observe that, in contrast to S&P's issuer ratings, our default probability anticipated much more quickly the imminent defaults of Enron, Parmalat, Delphi, and General Motors. For Enron, S&P did not change the rating until days before the corporation declared bankruptcy (Panel A). However, one of the reasons why the default probability for Enron based on the Merton model **M3** increased may be due to a falling business cycle and a corresponding general increase of default likelihoods. Since S&P is, by their own definition, through the cycle, one may expect that it does not anticipate such economic downturns, but it must anticipate the relative worsening of Enron's individual credit standing in comparison to other companies. However, even our through-the-cycle model **TTC** picks up the worsening of Enron's creditworthiness already at the end of 2000. For Parmalat, the story is similar (Panel B). Parmalat defaulted in December 2003 and was the biggest European default at the time, but the S&P rating for Parmalat at the end of November 2003 was still unchanged. The S&P record for the defaulting car parts supplier Delphi is a bit better (Panel C). However, the ζ-score **A1** shows a better performance than S&P. The ζ-score

A1 correctly generates low probabilities of default before the end of 2004 and then increased the default risk significantly during the last year of Delphi's life. Compared to Merton's model **M3**, the ζ-score **A1** did not produce false default signals during the performing years of Delphi. For General Motors (Panel D), S&P's rating correctly anticipated its financial distress. This is reflected by the fact that they gradually decreased their rating until General Motors' default. However, our alternative models also seem to perform well in the case of General Motors.

[Figure 6 about here]

While the focus in Figure 5 is on companies that actually defaulted, we may now want to switch focus to companies that have not defaulted. In Figure 6, we plot the default forecasts for Fiat, General Electric, Coca-Cola, and McDonald's. We observe that Merton's model **M3** may produce more false alarms than the other forecasting approaches. Especially for General Electric, Coca-Cola, and McDonald's, S&P's ratings and our adjusted default probability based on **TTC** are more stable over the business cycle than **M3**, as seen in Figure 6. Before recessionary periods, the probability of default based on models **M3**, **A1**, and **PIT** is by construction higher on average than before expansionary periods of the business cycle. This can be seen most prominently with GE and the difference between **PIT** and **TTC** after 09/11 and after the bankruptcy of Lehman Brothers in 09/08. Merton's model **M3** produces stronger false alarms in the form of spikes in the case of General Electric (Panel B) and McDonald's (Panel D) than our default forecasts from model **TTC**. The default probability based on **M3** created some significant spikes even though the two companies never defaulted on their debt. We also observe two spikes in the time series of Merton's default probability **M3** for Coca-Cola (Panel C). The S&P rating of Coca-Cola, however, remained constant. On the other hand, S&P was rather slow in adjusting the rating of Fiat (Panel A). With Fiat, the track record of S&P is rather dismal. In 2007, the default probability based on **TTC** for Fiat was at a historic low. Merton's **M3** default probability measures almost no default risk. S&P followed with some delay in 2008, but it is during that time that the credit outlook of Fiat deteriorated significantly due to the financial market crisis and the anticipated recessionary period. As a consequence, S&P reversed its erstwhile credit upgrade at the beginning of 2009, i.e., exactly at a time when the credit outlook of Fiat improved again.

From the inspection of these specific examples above, we may also argue that expressing the credit quality of a company in terms of a continuous variable instead of a discrete ordinal measure has some additional advantages from a stability viewpoint. We may argue that credit quality should be expressed as a continuous variable. When a discrete credit rating changes, the corresponding default probability changes discontinuously. This discontinuity may well spread over the markets, especially if regulatory capital is a function of these ratings, leading to a substantial drop in the company's share price.[12] A discontinuous change during difficult times may further nourish market uncertainty by giving an incentive for creditors to run, precipitating a collapse and contagion. Our default probabilities are continuous assessments of credit quality and may thus cause less market disruption when the credit quality of a company deteriorates than the change of a discrete rating. Furthermore, the through-the-cycle version **TTC** of our model is also a continuous credit assessment but by construction less volatile than our original point-in-time model **PIT** and therefore potentially even more appropriate for regulatory capital charges.

V. Out-of-Sample Validation

To test the out-of-sample forecasting performance of the estimated models in the previous section, we will consider test statistics for the validation sample. We want to make forecasts with the most powerful predictor and the resulting PD estimates of the most powerful predictor must be calibrated. For regulatory purposes, the best performing through-the-cycle approach might be preferable to the most powerful point-in-time model. Thus, in the following, we first discuss the results on discrimination and then on calibration.[13]

A. Discrimination Testing

To assess the discriminatory power of the various models, we construct the empirical equivalent of the Lorenz curve as defined in (1). That is, we order the observations according to their estimated default probabilities from lowest to highest. Then, we analyze the number of defaults included in a given fraction of the population. The results for the different models are shown in Table IV.

[TABLE IV about here]

We see, for instance, that 25.54% of all defaults are included in the first nine deciles of the S&P ratings. But for model **PIT**, only 12.60%, or one-half as many, defaults are included. Approximately seven-eighths of all defaults are found in the 10th decile with the highest PD estimates versus only three-quarters in the case of S&P. Even if, for fairness, we look at the through-the-cycle version of our model (**TTC**), we find that 17.72% of all defaulters are included in the first nine deciles. Therefore, we can conclude that the difference between S&P ratings and **PIT** indicates a superior discriminatory power of the latter model. The superiority for the 90% quantile even holds for **TTC**, i.e., under the provision for business cycle effects.

[Figure 7 about here]

In Figure 7, we plot the empirical Lorenz curves with which we can consider the discrimination for any quantile. The area above the curve serves as our summary statistic for a model's discriminatory power. The larger the area, the better the discriminatory power. The **PIT** model generates an area above the Lorenz curve of 94.83%, compared with an area of 92.34% for Altman's ζ-score **A1**, 93.77% for Merton's model **M3**, and 91.84% for S&P ratings. Even when we adjust our default probabilities for the business cycle, we still have a higher area above the Lorenz curve than we would obtain with S&P's credit ratings, i.e., 93.30% for **TTC** versus 91.84% for S&P. Thus, the better discriminatory power of our model **PIT** compared to the ratings is not purely based on the fact that our PD increases on average before recessionary periods and decreases before expansions.

[TABLE V about here]

Table V presents the inferential statistics for the area above the Lorenz curve for **PIT**, Altman's ζ-score **A1**, Merton's rank and power transformed model **M3**, S&P ratings, and our business cycle adjusted model **TTC** with the t-values for the null hypothesis that the two areas above the Lorenz curve are equal. We also include the naive forecast with no discriminatory power at all, which assumes that all companies share the same default probability. As expected, we see that the approaches are all significantly different from the naive forecasts. We also observe that our model's **PIT** forecasts are significantly superior to all other models at the 1% significance level. Hence, the null hypothesis of having the same discriminatory power must be rejected. The business cycle adjusted version model **TTC**, however, is not statistically superior to S&P ratings at a reasonable confidence level. Therefore, our through-the-cycle model **TTC** is at least as good as S&P in terms of discrimination ability. We can reach the same conclusion when we look at Panel B of Figure 8 for the monthly time series of areas above the Lorenz curves. For a clear majority of forecasting periods, model **TTC** outperforms the S&P ratings. Note that, for a single time period, the ranking of **PIT** and **TTC** is the same by construction and therefore both models have the same discrimination in Panel B of Figure 8.

In the literature, e.g., Chava and Jarrow (2004) and Hillegeist, Keating, Cram, and Lundstedt (2004), it has been concluded that the inclusion of market-based variables in the model renders the accounting variables relatively unimportant for default forecasting. If we look at the test statistics in Table V, we do not find support for such a conclusion. Although the accounting-based model **A1** fails to be significantly more powerful than the S&P ratings, its marginal contribution to **PIT** is highly significant. The marginal default signals stemming from accounting data are clearly relevant when combined with the distance to default measure, in that **PIT** significantly outperforms **M3**. We made the same observation when performing likelihood ratio tests in the estimation sample. Such an empirical finding is substantiated theoretically by Proposition 1 and is therefore not surprising.

It is further noteworthy that **TTC** is significantly less powerful than **PIT** even though **TTC** outperforms the S&P ratings. In light of these results, it seems questionable whether, as propounded by many, default forecasts should be credit-cycle adjusted. Such an adjustment eventually results in forecasts that are less powerful. Based on Lemma 1, we expected such an outcome, since the information set with cycle data is greater than the set without. From an econometric perspective, it is more difficult to include business cycle information in the forecasting than to simply filter it out. In any case, to avoid pro-cyclical regulatory bank capital charges, our **TTC** version provides non-cyclical conditional default probabilities that can be used for such purposes. However, from a purely forecasting perspective, we clearly favor **PIT** over **TTC**.

Altogether, we find **PIT** to be the most powerful model and we should use this predictor to make powerful default forecasts. However, we still need to test whether the PD estimates of **PIT** are calibrated, according to Definition 2. If yes, we are actually done, if no, we should look for a strictly rank preserving transformation to make the **PIT** predictor not only powerful but also calibrated. However, for the calculation of non-cyclical regulatory or economic capital requirements we should, as a final step, also check whether the less powerful models **TTC** and S&P provide at least calibrated forecasts.

B. Calibration Testing

Calibrated forecasts provide an unbiased estimate for the mean default rate and calibrated forecasts differentiate correctly between low and high default probabilities. Blöchlinger and Leippold (2011) introduce the formal concept of level calibration for the former property and the concept of shape calibration for the latter, and they show that a predictor that is both level- and shape-calibrated is calibrated as in Definition 2. In the following, we will apply the single-period test statistics on level and shape following Blöchlinger and Leippold (2011), as well as the multi-period summary statistics of Blöchlinger (2012). The global null hypothesis states that the default forecasts are calibrated in a given period. Thus, if we reject the null, we find miscalibrated forecasts in that period. The aggregation of a series of single-period statistics into one multi-period statistic solves the multiple testing problem.

[Figure 8 about here]

The level test compares realized with predicted default frequencies under the assumption of a default correlation. To this end, we plot in Panel A of Figure 8 the realized 1-year default rate, expressed as the fraction of firms that failed during the 1-year period. We compare this realized rate to the mean PD estimate that was made at the beginning of the 1-year period. We first observe that there is a strong variation over time in the frequency of defaults. In particular, there was a long lasting increase around 1998 and a significant spike during the last crisis, with default rates back to the pre-crisis level in 2011. Second, the S&P ratings mapped into "through the cycle default probabilities" are very stable over time. In quiet times, they significantly overestimate defaults while in turbulent periods, they heavily underestimate defaults. However, the S&P ratings are not intended to provide absolute measures of default risk, but our **PIT** model should. Hence, third, we observe that our model **PIT** captures much of the broad variation in corporate failures over time. It somewhat underestimates failures after 2000, and then overestimates slightly around 2010. However, remember that not a single parameter was re-estimated after October 2000 (the end of the in-sample period), i.e, the whole model remains completely unchanged after October 2000. By construction, our mean PD estimate of model **TTC** is constant over time.

[Figure 9 about here]

On the left (right) hand side of Figure 9, we compare the expected Lorenz curve with the empirical curve of **PIT** (**TTC**) to test the shape calibration hypothesis. The expected Lorenz curve shows the fraction of expected defaults among the $x\%$ observations with the lowest PD estimates. The empirical Lorenz curve then shows the corresponding fraction of realized defaults.[14] According to the PD ratios of **PIT**, we expect 1.3% (15.7%) of all defaults among the 50% (90%) observations with the lowest PD estimates. This computation gives us two points on the expected Lorenz curve. On the realized Lorenz curve, we obtain the corresponding points by replacing expected defaults with realized defaults. That is, we experienced 0.8% (12.6%) defaults among the 50% (90%) observations with the lowest PD estimates, which gives us two points on the empirical Lorenz curve. Repeating this procedure for all percentiles gives us the expected and the realized Lorenz curves. Under the null hypothesis of shape-calibrated forecasts, the expected curve and the true curve coincide. By inspection of Figure 9, we observe that the empirical curve and the expected curve are quite close both for **PIT** and **TTC**. However, we would still need a statistical test to claim that our models **PIT** and **TTC** are indeed calibrated.

[TABLE VI about here]

Therefore, we now apply the single period statistical tests of Blöchlinger and Leippold (2011), as well as the multi-period aggregation of Blöchlinger (2012). Loosely speaking, the shape test tells us whether in Figure 9, the empirical Lorenz curve is statistically different from the expected Lorenz curve. The level test tells us whether in Figure 8, the trajectories of the predicted and observed defaults are statistically different. In Table VI, we present the relevant test statistics, i.e., level, shape, and the combined statistic, for **PIT**. The level statistic compares the realized default rate with the mean PD estimate. The shape statistic depends on the difference between the realized and expected areas above the Lorenz curve. Both normalized statistics are χ^2-distributed with one degree of freedom and stochastically independent. The combined statistic is an aggregated test on both level and shape. It follows a χ^2-distribution with two degrees of freedom. In Table VI, we observe that in each year our model **PIT** provides calibrated forecasts.

Table VI shows the multi-period summary statistic on the last line. The multi-period statistic is the sum of the corresponding independent single period statistics and thus is χ^2-distributed. The number of degrees of freedom equals the number of time periods for the shape statistics and level statistics, and twice this number for the combined multi-period statistic. The null hypothesis of level- and shape-calibrated forecasts clearly cannot be rejected for our PD estimates obtained with **PIT**. Consequently, the **PIT** score in (16) represents the "true" distance to default measured as the number of standard deviations away from default, so that the PD estimates of **PIT** are indeed calibrated.

Table VII tabulates the multi-period level, shape, and combined statistics for **PIT** in comparison to our benchmark models **A1**, **M3**, S&P, and **TTC**. Apart from S&P, the other models seem to be calibrated overall as well, as seen from the p-values of the combined $\chi^2_{<22>}$-statistics in brackets. That is, the cycle adjusted version **TTC** of our point-in-time model provides calibrated forecasts even though these forecasts are based only on relative risk and therefore on a restricted information set that in effect disregards macroeconomic data. We can see that the two through-the-cycle approaches, **TTC** and S&P, struggle with the level calibration, as expected since default dependencies are fully driven by macroeconomic factors which can cause more tail events. The three other approaches implicitly incorporate the macroeconomic outlook when making default predictions so that only unanticipated macroeconomic forecasting errors drive default dependencies. Unexpectedly, however, our S&P PD function, unlike, **TTC** is clearly not shape calibrated.

The PD ratio between the S&P rating classes seems not to be stable over time, as seen in Table VIII. The relative ranking according to S&P in 2002 had a different meaning in terms of PD ratios between classes than in 2010. The PD ratio between non-investment grade issuers and investment grade borrowers was lower in 2002 than in 2010. We cannot accept the null hypothesis that the PD ratios are constant over time. As a consequence, we must clearly reject the calibration hypothesis for S&P ratings. As listed in Table VIII, the realized area above the Lorenz curve in the period from 10/2001 to 10/2002 is 86.5% with a much higher expectation of 91.8%, causing a $\chi^2_{<1>}$-value of 13.1 for that period. Hence, we observe far too many defaults among well rated borrowers relative to low rated obligors when benchmarked against S&P's PD ratios from Panel B of Figure 3. On the other hand, the expected area between 10/2009 and 10/2010 of 90.8% is much lower than the realized area of 97.3%, causing a $\chi^2_{<1>}$-value of 7.4. In that year, we observe too few defaults in good rating classes in comparison to low credit categories. Therefore, these two time periods with a $\chi^2_{<2>}$-value of 20.5 (=13.1+7.4) out of eleven periods already explain two-thirds of the $\chi^2_{<11>}$-multi-period shape statistic of 30.6, as shown in Table VII and Table VIII. This finding supports the conclusion of Alp (2013), that a structural shift occurs from 2003 onwards towards more stringent ratings as investment grade standards tightened.

Overall, we conclude this section with the observation that **PIT** is calibrated and is the most powerful model. It is therefore recommended for making default predictions. Model **TTC** is a calibrated and a cycle-adjusted version of **PIT** which is at least as powerful as the uncalibrated S&P ratings and is recommended as input for the calculation of non-cyclical regulatory capital requirements.

VI. Conclusion

We present a default prediction model that provides default probabilities for forecasting corporate credit defaults in a 1-year forecast horizon. Our model combines the distance-to-default model of Merton (1974) and the reduced-form statistical approach of Altman (1968). Within a structural model, we show that neither Merton's nor Altman's approach can lead to powerful default forecasts when there are information asymmetries between the firm's managers and capital market participants, with the effect that the firm's asset value is only observable with noise. In fact, both Merton's and Altman's approach are nested within our general structural model as two special cases when the asset value is either perfectly observable or unknown.

Our empirical analysis demonstrates that our default predictor provides calibrated forecasts in each state of a credit cycle, i.e., the expected default frequencies match closely the realized frequencies. We show that too restrictive distributional assumptions, such as normality, would result in a lack of fit between the expected and realized default frequencies, and we resort to non-parametric rank and power transformations to obtain well fitting models. In addition to being calibrated, our model better discriminates defaulting firms from non-defaulting firms than the stand-alone concepts and also better than Standard & Poor's issuer ratings. Our model fulfills the regulators' request that an alternative for agency ratings should be transparent and replicable.

We also present an alternative version of our model that is based only on the relative rankings of the current population of borrowers. This version in effect corrects our forecasts for business cycle influences. In this cyclically adjusted version, the realized mean default rate deviates more from the predicted level—as is to be expected when abstracting from the current state of the economy—but the realized ratios of default frequencies in different percentiles still match closely the expected ratios. Unlike our model, the interpretation of S&P rating classes in terms of probability ratios is not stable over time. S&P seems to have become more stringent in its rating assignments. In any case, even our cycle-adjusted version is calibrated and still at least as powerful as S&P ratings.

Especially for the sizeable money market, when both the counter-party default risk as well as the default risk of the underlying collateral must be assessed for forecast horizons of up to one year, our model is a sound and viable alternative to agency ratings. The adjusted version of our model filters out macroeconomic influences and is therefore non-cyclical and well suited for the computation of regulatory capital charges. We conclude that for the case of corporate default risk, appropriate, transparent, powerful, and calibrated alternatives to agency ratings do indeed exist and are readily available to be applied in practice.

References

Akaike, H., 1974, A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19, 716–723.

Alp, A., 2013, Structural shifts in credit rating standards, *Journal of Finance* 68, 2435–2470.

Altman, E.I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance* 23, 589–609.

——— , R.G. Haldeman, and P. Narayanan, 1977, ZETA analysis: A new model to identify bankruptcy risk of corporations, *Journal of Banking & Finance* 1, 29–54.

Altman, E. I., and H. A. Rijken, 2006, A point-in-time perspective on through-the-cycle ratings, *Financial Analysts Journal* 62, 54–70.

Association for Financial Professionals, 2002, *Rating agencies survey: Accuracy, Timeliness and Regulation*www.afponline.org.

Basel Committee on Banking Supervision, 2005, Studies on the validation of internal rating systems, Working Paper 14 Bank for International Settlement.

Bharath, S.T., and T. Shumway, 2004, Forecasting default with the KMV-Merton model, Discussion paper, AFA 2006 Boston Meetings Paper.

——— , 2008, Forecasting default with the Merton distance to default model, *Review of Financial Studies* 21, 1339–1369.

Blöchlinger, A., 2011, Arbitrage-free credit pricing using default probabilities and risk sensitivities, *Journal of Banking & Finance* pp. 268–281.

Blöchlinger, A., 2012, Validation of default probabilities, *Journal of Financial and Quantitative Analysis* 47, 1089–1123.

——— , and M. Leippold, 2011, A new goodness-of-fit test for event forecasting and its application to credit defaults, *Management Science* 57, 471–486.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance* 63, 2899–2939.

Chava, S., R. Ganduri, and C. Ornthanalai, 2012, Are Credit Ratings Still Relevant?, Working paper Georgia Institute of Technology.

Chava, Sudheer, and Robert A. Jarrow, 2004, Bankruptcy prediction with industry effects, *Review of Finance* 8, 537–569.

Conover, W. J., and R. L. Iman, 1981, Rank transformations as a bridge between parametric and nonparametric Statistics, *The American Statistician* 35, 124–129.

Coval, J.D., J.W. Jurek, and E. Stafford, 2009, Economic catastrophe bonds, *The American Economic Review* 99, 628–666.

Crosbie, P., and J. Bohn, 2003, Modeling default risk, *Moody's KMV*.

Duan, Jin-Chuan, Jie Sun, and Tao Wang, 2012, Multiperiod corporate default prediction—A forward intensity approach, *Journal of Econometrics* 170, 191–209.

Duffie, D., A. Eckner, G. Horel, and L. Saita, 2009, Frailty correlated default, *Journal of Finance* 64, 2089–2123.

Duffie, D., and D. Lando, 2001, Term structures of credit spreads with incomplete accounting information, *Econometrica* 69, 633–664.

Duffie, D., L. Saita, and K. Wang, 2007, Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics* 83, 635–665.

Fink, R., 2003, Where credit is due, *CFO Magazine* November.

Fischer, Edwin O., Robert Heinkel, and Josef Zechner, 1989, Dynamic recapitalization policies and the role of call premia and issue discounts, *Journal of Financial and Quantitative Analysis* 24, 427–446.

Hamilton, J. D., 1994, *Time Series Analysis* (Princeton University Press).

Hillegeist, S., E. Keating, D. Cram, and K. Lundstedt, 2004, Assessing the probability of bankruptcy, *Review of Accounting Studies* 9, 5–34.

Hilscher, J., and M. I. Wilson, 2012, Credit ratings and credit risk, Working paper Brandeis University and Oxford University.

Hosmer, D. W., T. Hosmer, S. le Cessie, and S. Lemeshow, 1997, A comparison of goodness-of-fit tests for the logistic regression model, *Statistics in Medicine* 16, 965–980.

Kane, Gregory D., and Nancy L. Meade, 1998, Ratio analysis using rank transformation, *Review of Quantitative Finance and Accounting* 10, 59–74.

Kane, Gregory D., Frederick M. Richardson, and Nancy L. Meade, 1998, Rank transformations and the prediction of corporate failure, *Contemporary Accounting Research* 15, 145–166.

Leland, Hayne E, 1994, Corporate debt value, bond covenants, and optimal capital structure, *Journal of Finance* 49, 1213–52.

Lorenz, M. O., 1905, Methods of measuring the concentration of wealth, *Publications of the American Statistical Association* 9, 209–219.

Maddala, G. S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics* (Cambridge University Press).

Merton, R.C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–470.

Shumway, T., 2001, Forecasting bankruptcy more accurately: A simple hazard model, *The Journal of Business* 74, 101–24.

Spearman, C., 1904, The proof and measurement of association between two things, *American Journal of Psychology* 15, 88–103.

Vassalou, M., and Y. Xing, 2004, Default risk in equity returns, *The Journal of Finance* 59, 831–868.

Appendix A. Proofs

Proof of Lemma 1. For readability, we define $P = \mathbb{E}[Y|\mathcal{G}]$. We have to show that the Lorenz curve of P lies below the Lorenz curve of any $\mathcal{G}$-measurable variable S. Without loss of generality but for the sake of clarity, we assume that both P and S are continuous random variables.

According to the definition of the Lorenz curve, P is more powerful than S if

$$\mathbb{P}\{S \leq s|Y = 1\} \geq \mathbb{P}\{P \leq p|Y = 1\},$$

for all $s \in (-\infty, +\infty)$ and corresponding $p \in [0, 1]$ such that $\mathbb{P}\{S \leq s\} = \mathbb{P}\{P \leq p\}$. To facilitate the notation, we define the events $\mathbb{A}(s)$ and $\mathbb{B}(p)$ as

$$\mathbb{A}(s) = \{S \leq s\} \text{ and } \mathbb{B}(p) = \{P \leq p\}, \tag{A1}$$

where for arbitrary $s \in (-\infty, \infty)$, $p \in [0, 1]$ is chosen so that $\mathbb{P}\{\mathbb{A}\} = \mathbb{P}\{\mathbb{B}\}$. Due to the following decomposition:[15]

$$\mathbb{A} = \mathbb{A} \cap \mathbb{B} + \mathbb{A} \cap \bar{\mathbb{B}} \text{ and } \mathbb{B} = \mathbb{A} \cap \mathbb{B} + \bar{\mathbb{A}} \cap \mathbb{B},$$

we deduce that $\mathbb{P}\left\{\mathbb{A} \cap \bar{\mathbb{B}}\right\} = \mathbb{P}\left\{\bar{\mathbb{A}} \cap \mathbb{B}\right\}$. We distinguish between two cases: $\mathbb{P}\left\{\mathbb{A} \cap \bar{\mathbb{B}}\right\} = \mathbb{P}\left\{\bar{\mathbb{A}} \cap \mathbb{B}\right\} = 0$ and $\mathbb{P}\left\{\mathbb{A} \cap \bar{\mathbb{B}}\right\} = \mathbb{P}\left\{\bar{\mathbb{A}} \cap \mathbb{B}\right\} > 0$. In the first case, we can readily deduce $\mathbb{A} = \mathbb{B}$, up to a set of probability measure zero, and therefore

$$\mathbb{P}\{\mathbb{A}|Y = 1\} = \mathbb{P}\{\mathbb{B}|Y = 1\}. \tag{A2}$$

In the second case, if $\mathbb{P}\left\{\mathbb{A} \cap \bar{\mathbb{B}}\right\} = \mathbb{P}\left\{\bar{\mathbb{A}} \cap \mathbb{B}\right\} > 0$, we have

$$\mathbb{E}\left[Y|\mathbb{A} \cap \bar{\mathbb{B}}\right] = \mathbb{E}\left[P|\mathbb{A} \cap \bar{\mathbb{B}}\right] > \mathbb{E}\left[P|\bar{\mathbb{A}} \cap \mathbb{B}\right] = \mathbb{E}\left[Y|\bar{\mathbb{A}} \cap \mathbb{B}\right],$$

where the equalities follow by the law of iterated expectations. The inequality follows from the fact that $\mathbb{E}\left[P|\bar{\mathbb{B}}\right] > \mathbb{E}[P|\mathbb{B}]$ and $\mathbb{A}, \bar{\mathbb{A}} \in \mathcal{G}$. We multiply both sides by $\mathbb{P}\left\{\mathbb{A} \cap \bar{\mathbb{B}}\right\}$ $(= \mathbb{P}\left\{\bar{\mathbb{A}} \cap \mathbb{B}\right\})$:

$$\mathbb{E}\left[\mathbf{1}_{\mathbb{A}}\mathbf{1}_{\bar{\mathbb{B}}}Y\right] = \mathbb{E}\left[\mathbf{1}_{\mathbb{A}}\mathbf{1}_{\bar{\mathbb{B}}}P\right] > \mathbb{E}\left[\mathbf{1}_{\bar{\mathbb{A}}}\mathbf{1}_{\mathbb{B}}P\right] = \mathbb{E}\left[\mathbf{1}_{\bar{\mathbb{A}}}\mathbf{1}_{\mathbb{B}}Y\right],$$

where $\mathbf{1}_{\mathbb{A}}$ is the indicator variable for the event $\mathbb{A}$. Adding $\mathbb{E}\left[\mathbf{1}_{\mathbb{A}}\mathbf{1}_{\mathbb{B}}Y\right]$ $(= \mathbb{E}\left[\mathbf{1}_{\mathbb{A}}\mathbf{1}_{\mathbb{B}}P\right])$ to both sides, we obtain

$$\mathbb{E}\left[\mathbf{1}_{\mathbb{A}}Y\right] = \mathbb{E}\left[\mathbf{1}_{\mathbb{A}}P\right] > \mathbb{E}\left[\mathbf{1}_{\mathbb{B}}P\right] = \mathbb{E}\left[\mathbf{1}_{\mathbb{B}}Y\right].$$

Dividing both sides by $\mathbb{P}\left\{Y = 1\right\}$ results in

$$\mathbb{P}\left\{\mathbb{A}|Y = 1\right\} > \mathbb{P}\left\{\mathbb{B}|Y = 1\right\}. \tag{A3}$$

Combining (A2) and (A3), we obtain:

$$\mathbb{P}\left\{\mathbb{A}|Y = 1\right\} \geq \mathbb{P}\left\{\mathbb{B}|Y = 1\right\} \text{ whenever } \mathbb{P}\left\{\mathbb{A}\right\} = \mathbb{P}\left\{\mathbb{B}\right\},$$

for all $\mathbb{A}$ and $\mathbb{B}$ as defined in (A1). □

Proof of Proposition 1. By Assumption 2, the following linear transformation is Gaussian:

$$\left.\begin{pmatrix} \log V^* + U \\ \log V^* + r \end{pmatrix}\right| X \sim N\left(\begin{pmatrix} \gamma^\top X - \alpha^2/2 \\ (\gamma + \lambda)^\top X - \sigma^2/2 \end{pmatrix}, \begin{pmatrix} \delta^2 + \alpha^2 & \delta^2 \\ \delta^2 & \delta^2 + \sigma^2 \end{pmatrix}\right).$$

The conditional distribution of $\log V^* + r|\log V^* + U, X$ is therefore also Gaussian (see, e.g., Hamilton (1994), p. 102 for a proof):

$$\log V^* + r|\, V, X \sim N\left((\gamma + \lambda)^\top X - \frac{\sigma^2}{2} + \frac{\delta^2}{\delta^2 + \alpha^2}\left(\log V - \gamma^\top X + \frac{1}{2}\alpha^2\right), \delta^2 + \sigma^2 - \frac{\delta^4}{\delta^2 + \alpha^2}\right),$$

where $\log V = \log V^* + U$ from Assumption 1. Hence, the conditional expectation of the default indicator $\mathbf{1}_{\{V^* \exp(r) \leq F\}}$ given F, V, and X can be written as

$$\mathbb{P}\left\{V^* \exp(r) \leq F|\, F, V, X\right\} = \Phi\left(\frac{\log F - (\gamma + \lambda)^\top X + \frac{\sigma^2}{2} - \frac{\delta^2}{\delta^2 + \alpha^2}\left(\log V - \gamma^\top X + \frac{\alpha^2}{2}\right)}{\sqrt{\delta^2 + \sigma^2 - \frac{\delta^4}{\delta^2 + \alpha^2}}}\right). \tag{A4}$$

Recalling the definition of the distance to default measure DD,

$$DD = \frac{\log V - \log F + \lambda^\top X - 0.5\sigma^2}{\sigma},$$

we can write the default probability in (A4) as

$$\text{PD} = \Phi\left(\beta_0 + \beta_1 \log F + \beta_2^\top X + \beta_{DD}\, DD\right), \tag{A5}$$

with

$$\beta_{DD} = -\sigma \frac{1}{\sqrt{\sigma^2 + \delta^2 - \frac{\delta^4}{\delta^2 + \alpha^2}}} \frac{\delta^2}{\delta^2 + \alpha^2},$$

as well as

$$\beta_0 = \frac{1}{2}\frac{1}{\sqrt{\sigma^2+\delta^2-\frac{\delta^4}{\delta^2+\alpha^2}}}\frac{\alpha^2}{\delta^2+\alpha^2}\left[\sigma^2-\delta^2\right]$$

$$\beta_1 = \frac{1}{\sqrt{\sigma^2+\delta^2-\frac{\delta^4}{\delta^2+\alpha^2}}}\frac{\alpha^2}{\delta^2+\alpha^2}$$

$$\beta_2 = -\frac{1}{\sqrt{\sigma^2+\delta^2-\frac{\delta^4}{\delta^2+\alpha^2}}}\frac{\alpha^2}{\delta^2+\alpha^2}\left(\gamma+\lambda\right). \tag{A6}$$

Finally, when V^* is also known, then the only randomness stems from the return r, which is conditionally Gaussian with mean $\lambda^\top X - \sigma^2/2$ and variance σ^2 such that $\mathbb{P}\left\{V^* \exp(r) \leq F \mid F, V, X, V^*\right\} = \Phi\left(-DD^*\right)$. ☐

Proof of Proposition 2. We will first prove this for the conditional default probability given DD, i.e., Merton's adjusted PD in (11). By bivariate normality,

$$\begin{pmatrix} Z \\ DD \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_Z \\ \mu_{DD} \end{pmatrix}, \begin{pmatrix} \sigma_Z^2 & \rho\sigma_Z\sigma_{DD} \\ \rho\sigma_Z\sigma_{DD} & \sigma_{DD}^2 \end{pmatrix}\right), \tag{A7}$$

we have the following conditional distribution for Z given DD (see p. 102 in Hamilton (1994)):

$$Z|DD \sim N\left(\mu_Z + \rho\frac{\sigma_Z}{\sigma_{DD}}\left(DD-\mu_{DD}\right), \sigma_Z^2\left(1-\rho^2\right)\right). \tag{A8}$$

Using (A8) we can explicitly compute the conditional probability $\mathbb{P}\left\{Y=1 \mid DD\right\}$. First, we apply the law of iterated expectation:

$$\mathbb{P}\left\{Y=1 \mid DD\right\} = \mathbb{E}\left[\mathbb{P}\left\{Y=1 \mid Z, DD\right\} \mid DD\right] = \mathbb{E}\left[\Phi\left(Z+\beta_{DD}DD\right) \mid DD\right],$$

where the second equality imputes the conditional probability given Z and DD from (7). Next, we use the conditional normal distribution from (A8):

$$\begin{aligned}\mathbb{E}\left[\Phi\left(Z+\beta_{DD}DD\right) \mid DD\right] &= \frac{1}{\sigma_Z\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\Phi\left(\nu+\beta_{DD}DD\right)\phi\left(\frac{\nu-\mu_z-\rho\frac{\sigma_Z}{\sigma_{DD}}\left(DD-\mu_{DD}\right)}{\sigma_Z\sqrt{1-\rho^2}}\right)d\nu \\ &= \int_{-\infty}^{\infty}\Phi\left(\xi\,\sigma_Z\sqrt{1-\rho^2}+\mu_Z-\rho\frac{\sigma_Z}{\sigma_{DD}}\mu_{DD}+DD\left(\beta_{DD}+\rho\frac{\sigma_Z}{\sigma_{DD}}\right)\right)\phi(\xi)\,d\xi \\ &= \Phi\left(\frac{\mu_Z-\rho\frac{\sigma_Z}{\sigma_{DD}}\mu_{DD}+DD\left(\beta_{DD}+\rho\frac{\sigma_Z}{\sigma_{DD}}\right)}{\sqrt{1+\sigma_Z^2\left(1-\rho^2\right)}}\right),\end{aligned} \tag{A9}$$

where $\phi(\cdot)$ is the probability distribution function of a standard Gaussian variable. The second line makes a variable substitution, the last line computes the definite integral of a Gaussian function.

To calculate Altman's PD in (12), we start with the conditional random variable $DD|Z$:

$$DD|Z \sim N\left(\mu_{DD} + \rho \frac{\sigma_{DD}}{\sigma_Z}(Z - \mu_Z), \sigma_{DD}^2\left(1 - \rho^2\right)\right),$$

and perform exactly the same steps as above to obtain the desired result:

$$\mathbb{E}\left[\Phi\left(Z + \beta_{DD} DD\right) \middle| Z\right] = \Phi\left(\frac{\beta_{DD}\mu_{DD} - \rho\beta_{DD}\frac{\sigma_{DD}}{\sigma_Z}\mu_Z + Z\left(1 + \rho\beta_{DD}\frac{\sigma_{DD}}{\sigma_Z}\right)}{\sqrt{1 + \beta_{DD}^2\sigma_{DD}^2(1 - \rho^2)}}\right). \tag{A10}$$

Finally, to obtain the naive PD in (13), we either integrate over $DD \sim N(\mu_{DD}, \sigma_{DD}^2)$ in (A9), or over $Z \sim N(\mu_Z, \sigma_Z^2)$ in (A10), or directly over $Z + \beta_{DD} D \sim N(\mu_Z + \beta_{DD}\mu_{DD}, \sigma_Z^2 + \beta_{DD}^2\sigma_{DD}^2 + 2\rho\sigma_Z\beta_{DD}\sigma_{DD})$. All ways lead to the same naive PD in (13), by the law of iterated expectations,

$$\mathbb{E}\left[\mathbb{E}\left[\Phi\left(Z + \beta_{DD} DD\right) \middle| Z\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\Phi\left(Z + \beta_{DD} DD\right) \middle| DD\right]\right] = \mathbb{E}\left[\Phi\left(Z + \beta_{DD} DD\right)\right] = \mathbb{P}\left\{Y = 1\right\},$$

and by calculation of definite integrals of Gaussian functions. □

Proof of Proposition 3. By assumption, $\log F$ and X follow a multivariate Gaussian distribution:

$$\begin{pmatrix} \log F \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_F \\ \mu_X \end{pmatrix}, \begin{pmatrix} \sigma_F^2 & \Sigma_{F,X} \\ \Sigma_{F,X}^\top & \Sigma_{X,X} \end{pmatrix}\right), \tag{A11}$$

so that the transformations $Z = \beta_0 + \beta_1 \log F + \beta_2^\top X$, $DD = \left(\log(V/F) + \mu - 0.5\sigma^2\right)/\sigma$ follow the bivariate Gaussian distribution in (A7) with the parameters

$$\begin{aligned}
\mu_Z &= \beta_0 + \beta_1\mu_F + \beta_2^\top \mu_X, \; \sigma_Z^2 = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^\top \begin{pmatrix} \sigma_F^2 & \Sigma_{F,X} \\ \Sigma_{F,X}^\top & \Sigma_{X,X} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \\
\mu_{DD} &= \frac{1}{\sigma}\left((\lambda + \gamma)^\top \mu_X - \frac{1}{2}\alpha^2 - \mu_F - \frac{1}{2}\sigma^2\right), \\
\sigma_{DD}^2 &= \frac{1}{\sigma^2}\left(\begin{pmatrix} -1 \\ \lambda + \gamma \end{pmatrix}^\top \begin{pmatrix} \sigma_F^2 & \Sigma_{F,X} \\ \Sigma_{F,X}^\top & \Sigma_{X,X} \end{pmatrix} \begin{pmatrix} -1 \\ \lambda + \gamma \end{pmatrix} + \delta^2 + \alpha^2\right), \\
\sigma_{Z,DD} &= \frac{1}{\sigma^2}\begin{pmatrix} -1 \\ \lambda + \gamma \end{pmatrix}^\top \begin{pmatrix} \sigma_F^2 & \Sigma_{F,X} \\ \Sigma_{F,X}^\top & \Sigma_{X,X} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \; \rho = \frac{\sigma_{Z,DD}}{\sigma_Z \sigma_{DD}}.
\end{aligned}$$

From Proposition 2, we have that the conditional probability $\mathbb{P}\{Y=1|\,DD\} = \Phi\left(b_0 + b_{DD}DD\right)$ with b_0 and b_{DD} given by

$$b_0 = \frac{\mu_Z - \rho\mu_{DD}\frac{\sigma_Z}{\sigma_{DD}}}{\sqrt{1+\sigma_Z^2(1-\rho^2)}}, \text{ and } b_{DD} = \frac{\beta_{DD} + \rho\frac{\sigma_Z}{\sigma_{DD}}}{\sqrt{1+\sigma_Z^2(1-\rho^2)}}.$$

Now, using the definition of ω,

$$\omega := \begin{pmatrix} -1 \\ \gamma+\lambda \end{pmatrix}^\top \begin{pmatrix} \sigma_F^2 & \Sigma_{F,X} \\ \Sigma_{F,X}^\top & \Sigma_{X,X} \end{pmatrix} \begin{pmatrix} -1 \\ \gamma+\lambda \end{pmatrix} \geq 0,$$

together with the parameters β_0, β_1, β_2 given in (A6) from Proposition 1, we can write b_0 and b_{DD} with the original parameters of the multivariate distribution in (A11) as

$$b_0 = \frac{\frac{\alpha^2}{\alpha^2+\delta^2}\left[\frac{\sigma^2-\delta^2}{2} + \mu_F - (\gamma+\lambda)^\top\mu_X + \frac{\omega}{\omega+\delta^2+\alpha^2}\left\{(\gamma+\lambda)^\top\mu_X - \mu_F - \frac{\alpha^2+\sigma^2}{2}\right\}\right]}{\sqrt{\sigma^2+\delta^2-\frac{\delta^4}{\alpha^2+\delta^2}+\omega\left(\frac{\alpha^2}{\alpha^2+\delta^2}\right)^2\left(1-\frac{\omega}{\omega+\delta^2+\alpha^2}\right)}}, \tag{A12}$$

and

$$b_{DD} = -\frac{\sigma\left(\frac{\delta^2}{\alpha^2+\delta^2}+\frac{\alpha^2}{\alpha^2+\delta^2}\frac{\omega}{\omega+\delta^2+\alpha^2}\right)}{\sqrt{\sigma^2+\delta^2-\frac{\delta^4}{\alpha^2+\delta^2}+\omega\left(\frac{\alpha^2}{\alpha^2+\delta^2}\right)^2\left(1-\frac{\omega}{\omega+\delta^2+\alpha^2}\right)}},$$

with b_{DD} a continuous function of α. Further, b_{DD} is equal to -1 if $\alpha=0$, and $b_{DD}=0$ if $\alpha\to\infty$. Also, b_{DD} is strictly increasing in α^2, so that $\partial b_{DD}/\partial\alpha^2 > 0$, since

$$\frac{\partial b_{DD}}{\partial\alpha^2} = \frac{\frac{\sigma}{2}\left(\delta^2+\omega\right)\left[2\alpha^2\left(\delta^2+\sigma^2+\omega\right)+\left(\delta^2+\omega\right)\left(\delta^2+2\sigma^2+\omega\right)\right]}{\left(\alpha^2+\delta^2+\omega\right)^2\left[\alpha^2\left(\delta^2+\sigma^2+\omega\right)+\sigma^2\left(\delta^2+\omega\right)\right]\sqrt{\frac{\alpha^2\delta^2}{\alpha^2+\delta^2}+\sigma^2}\sqrt{\frac{\alpha^4\omega}{(\alpha^2(\delta^2+\sigma^2)+\delta^2\sigma^2)(\alpha^2+\delta^2+\omega)}+1}}.$$

Therefore, if $\alpha\in(0,\infty)$, then we have $b_{DD}\in(-1,0)$ by monotonicity, and from (A12) we see that b_0 is finite, i.e., $b_0\in(-\infty,+\infty)$. We determine the intersection point c for which Merton's original PD, $\Phi\left(-DD\right)$, coincides with the adjusted PD, $\Phi\left(b_0+b_{DD}DD\right)$:

$$\Phi\left(b_0+b_{DD}c\right) = \Phi\left(-c\right) \quad\Leftrightarrow\quad c = -\frac{b_0}{1+b_{DD}},$$

and we have therefore the following relation:

$$\begin{aligned} b_0 + b_{DD}DD &> -DD \quad \text{if } DD > c \\ b_0 + b_{DD}DD &< -DD \quad \text{if } DD < c. \end{aligned}$$

Since the cumulative distribution function $\Phi(\cdot)$ is monotonic we are done, i.e., with $\Phi(-DD)$ we overestimate the true default probability if $DD < c$, i.e., high risks, and we underestimate the true probability if $DD > c$, i.e., low risks. □

Appendix B. Replication of the Augmented Prediction Model

Appendix B is added so that our augmented prediction model **PIT** in (14) is fully transparent and easily replicable. We first show the computation of the distance to default measure. Second, we describe the calculation of Altman's financial ratios.

Constructing the Distance to Default Measure. The structural default model of Merton (1974) stipulates that the equity value of a firm can be interpreted as an option on the firm's asset. Under the assumption of a Gaussian distributed asset return, the equity value E of a firm having issued just one discount bond with a remaining maturity of one year satisfies

$$E = V\Phi(d+\sigma) - e^{-r_f} F\Phi(d), \tag{B1}$$

where d is the so-called risk-neutral distance to default measure

$$d = \frac{\log V - \log F + r_f - 0.5\sigma^2}{\sigma}. \tag{B2}$$

In Equation (B1), we denote by V the market value of the firm's total assets, F is the face amount of the firm's debt, r_f is the continuously compounded risk-free interest rate, $\Phi(\cdot)$ is the cumulative distribution function of a standardized Gaussian variable, and σ is the volatility of the firm's total assets. Merton's distance to default measure DD as given in (2) can be written as the risk-neutral distance to default measure d plus a standardized risk premium:

$$DD = d + \frac{\mu - r_f}{\sigma}, \tag{B3}$$

where μ is the expected return on the firm's total assets V. By Itô's lemma, asset volatility σ and equity volatility σ_E are related by

$$\sigma_E = \frac{V}{E}\frac{\partial E}{\partial V}\sigma = \frac{V}{E}\Phi(d+\sigma)\sigma, \tag{B4}$$

where d is defined in (B2).

We follow closely the implementation of Bharath and Shumway (2008) to estimate DD in (B3). The procedure can be described in four steps. In the first step, we estimate the equity volatility σ_E from historical daily stock returns over a rolling window of one year. In the second step, we choose a measure for the firm's nominal amount of debt F by using current liabilities plus one-half of long term debt. The third step is to collect daily market equity values and risk-free rates in the firm's home country. The equity value E unlike V is observable in the marketplace by multiplying the firm's outstanding shares by its current stock price. We approximate the 1-year risk-free rate r_f by the generic on-the-run government bill, note, or bond index.[16] The first three steps give us values for each of the variables in Equations (B1) and (B4) except for V and σ. Hence, in the fourth and most significant step, we estimate V and σ. Simultaneously solving the two Equations (B1) and (B4) for V and σ seems quite straightforward. However, Crosbie and Bohn (2003) state that "in practice the market leverage [F/V] moves around far too much for [Equation (B4)] to provide reasonable results." To resolve this issue, we first choose a starting value $\sigma = \sigma_E E/(E + F)$ and use this starting value together with Equation (B1) to infer the market value of each firm's assets every day for the previous year. We then compute the log differences from the implied asset values and use that time series to calculate standard deviation and mean to update our estimates for σ and μ. We iterate on σ in this way until convergence (i.e., until the absolute difference in adjacent σs is less than 0.001). With this last step we obtain V and σ to compute the risk-neutral distance to default d in (B2) as well as μ to compute the risk adjustment for DD in Equation (B3). For a small number of companies at certain points in time, the iterative procedure does not converge within 14 iterations. We drop the few non-converging observations.

Note that with this procedure, the estimate for μ can be lower than r_f. Based on our theoretical arguments, we have to work with the risk-adjusted DD as a function of μ instead of the risk-neutral d as a function of r_f. The estimation of μ is quite important as reported by Bharath and Shumway (2004). As somehow expected by our structural model, the default predictor that sets μ equal to the risk free rate r_f performs substantially worse in the empirical analysis of Bharath and Shumway (2004).

Altman's Accounting-Based Variables. Altman (1968)) and Altman, Haldeman, and Narayanan (1977) present two accounting-based credit scores. The former is called the Z-score with five financial ratios, the latter is the ζ-score with seven financial figures.

The Z-score incorporates the following five variables:

Z_1 compares liquid assets to total assets. The ratio is computed as working capital (current assets minus current liabilities) divided by total assets.

Z_2 quantifies the cumulative profitability over time as the ratio of retained earnings to total assets.

Z_3 illustrates the productivity of the firm's assets, abstracting from leverage and taxes. The ratio is calculated as earnings before interest and taxes (EBIT) divided by total assets. We compute EBIT by summing up EBIT for the most recent four quarters. EBIT is after amortization of goodwill.

Z_4 is a measure for leverage, and the only ratio of the Z-score which contains market prices. The ratio is defined as the market value of equity divided by the book value of total debt. We compute the market value of the equity by multiplying the number of shares by the closing price of the share. The number of shares is the number of primary common shares of all classes outstanding, net of treasury shares.

Z_5 quantifies the sales-generating ability of the firm as the ratio of sales to total assets. We consider the trailing twelve-month sales, i.e., the sum of the quarterly sales of the most recent four quarters.

The ζ-score model contains the following seven variables:

ζ_1 corresponds to the ratio Z_3 of the older Z-score model, EBIT divided by total assets. However, original ζ-score model Altman, Haldeman, and Narayanan (1977) includes total tangible assets as the denominator. Since many companies in our sample do not disclose intangibles, we use total assets instead.

ζ_2 quantifies the stability of the earnings. Altman, Haldeman, and Narayanan (1977) describe this measure as a normalized measure of the standard error of the estimate around a ten-year trend in ζ_1. We compute this variable as the coefficient of determination (R-squared) with the trailing twelve-month EBIT as the dependent variable, and time as the independent variable. If EBIT follows a trend, this variable is close to one, whereas it is close to zero if the time series of EBIT cannot be well explained with a trend line. We choose a five-year trend. We compute R-squared as soon as three values for ζ_1 are available (this is usually the sixth reporting quarter).

ζ_3 measures the ability of the firm to meet its liabilities (debt service). This variable is computed as EBIT divided by total interest payments, plus working capital divided by long term debt. Total interest payments are computed as the trailing twelve-month payments, i.e., the cumulated interest payments of the latest four quarters. Long term debt is total liabilities minus current liabilities.

ζ_4 is Z_2 of the Z-score, cumulative profitability, computed as retained earnings divided by total assets.

ζ_5 measures liquidity by the current ratio, which is current assets divided by current liabilities.

ζ_6 is the capitalization, measured as common equity divided by total capital. In the numerator of this ratio, the common equity is measured by the market value of the total equity, rather than the book value.

ζ_7 measures size by total assets relative to the median asset value among the underlying population of issuers. Otherwise, size would be the only variable in the Z-score and the ζ-score model which has a unit, and from an econometric perspective it would have a unit root, and therefore not be mean-reverting.

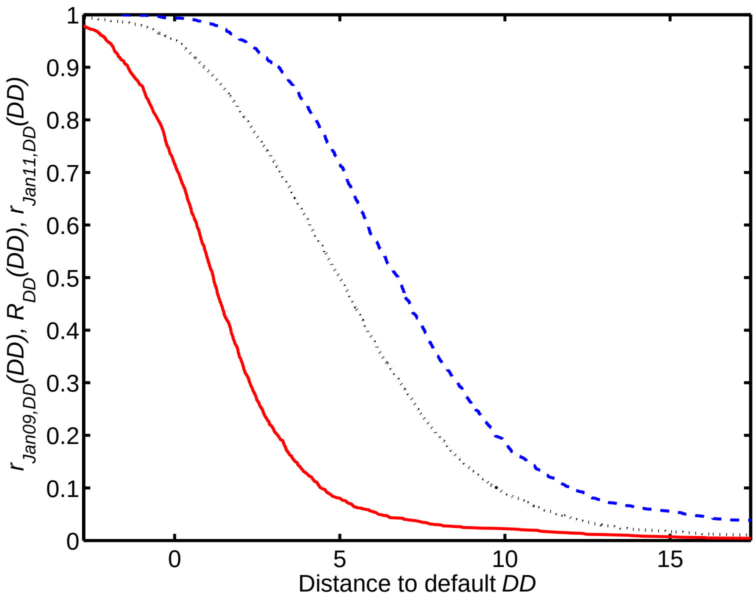

Figure 1. Long-term R_{DD} and Current Ranking $r_{t,DD}$ of Distance to Default DD
We applied a rank transformation R_{DD} (dotted line) to Merton's distance to default measure DD for a population of issuers over a business cycle (=long-term ranking) and we applied a rank transformation $r_{t,DD}$ to all observations at the end of January 2009 (solid line) and the end of January 2011 (dashed line).

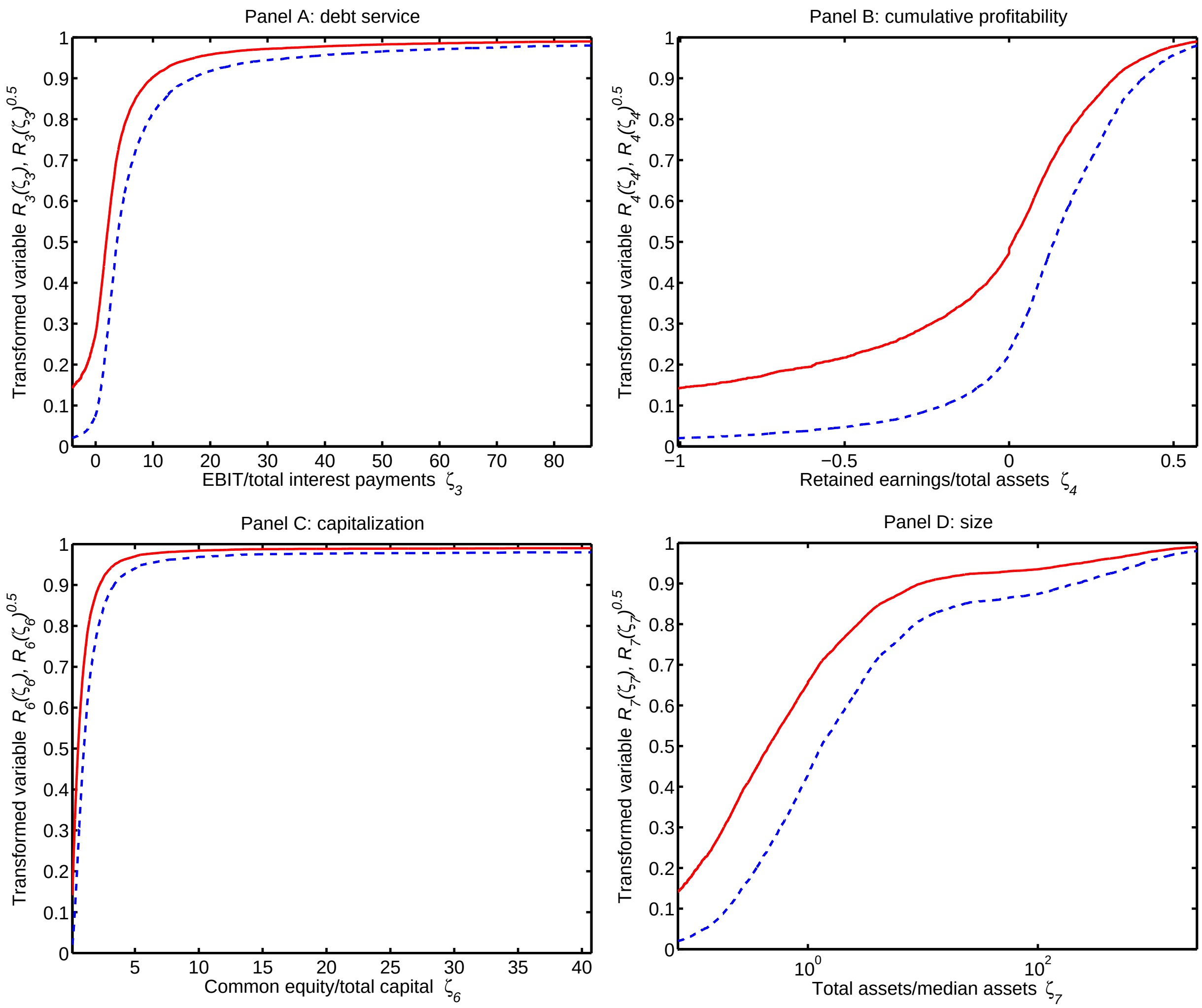

Figure 2. Rank and Square Root Transformation of Financial Ratios
We applied a rank transformation (dashed line) to the original financial ratios suggested by Altman, Haldeman, and Narayanan (1977) to predict corporate defaults. To obtain a better fit of the data, we further applied a square root transformation (solid line) to the rank transform of the four ζ variables. The four functions are estimated on an estimation sample and then applied to a validation sample. We modified ζ_7 from Altman, Haldeman, and Narayanan (1977) in that we measure the size not in dollars but relatively to the median total assets of the underlying population of borrowers at that time. Otherwise, the time series of size would contain a unit root.

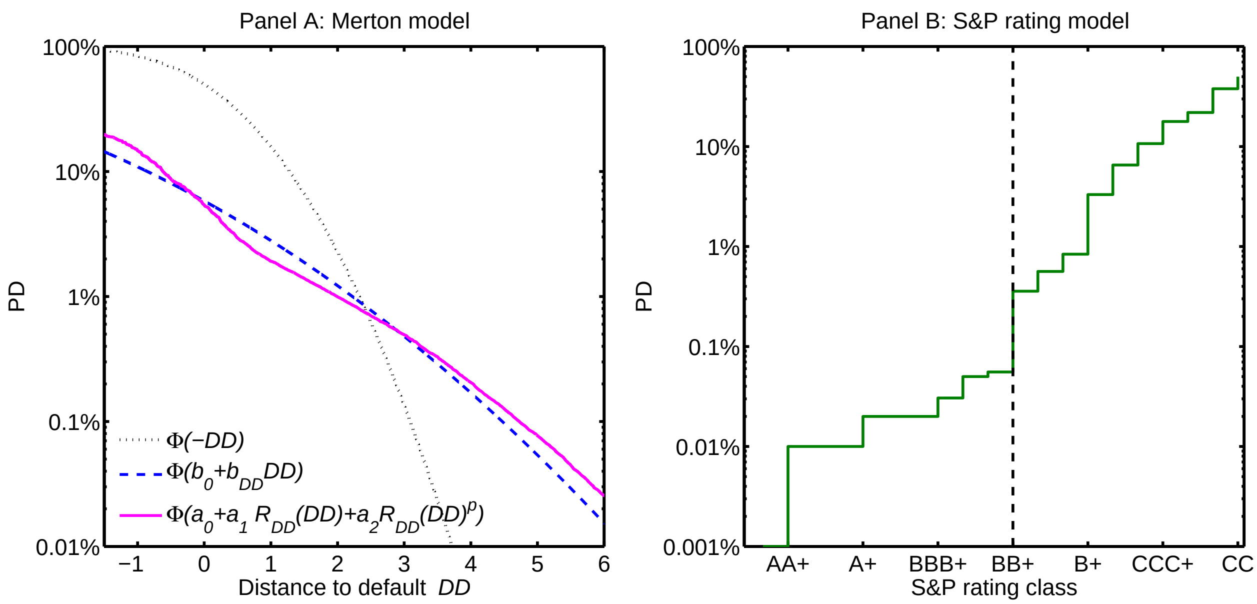

Figure 3. Mapping Distance to Default and Ratings to Default Probabilities
In Panel A we plot the mapping of the distance to default to default probabilities using different Merton models: the original Merton model **M1** (dotted line), Merton's adjusted model **M2** (dashed line), and Merton's rank- and power-transformed model **M3** (solid line). Panel B shows the mapping of S&P rating classes into default probabilities estimated by simple count statistics. To have strictly positive PD values and to have a monotone PD function, we interpolate between BBB+ and BBB- to obtain the PD for BBB rated borrowers and we assign 0.02% to A+, A, A-, 0.01% to AA+, AA, AA- as well as 0.001% to AAA. The dashed vertical line separates investment from non-investment grades.

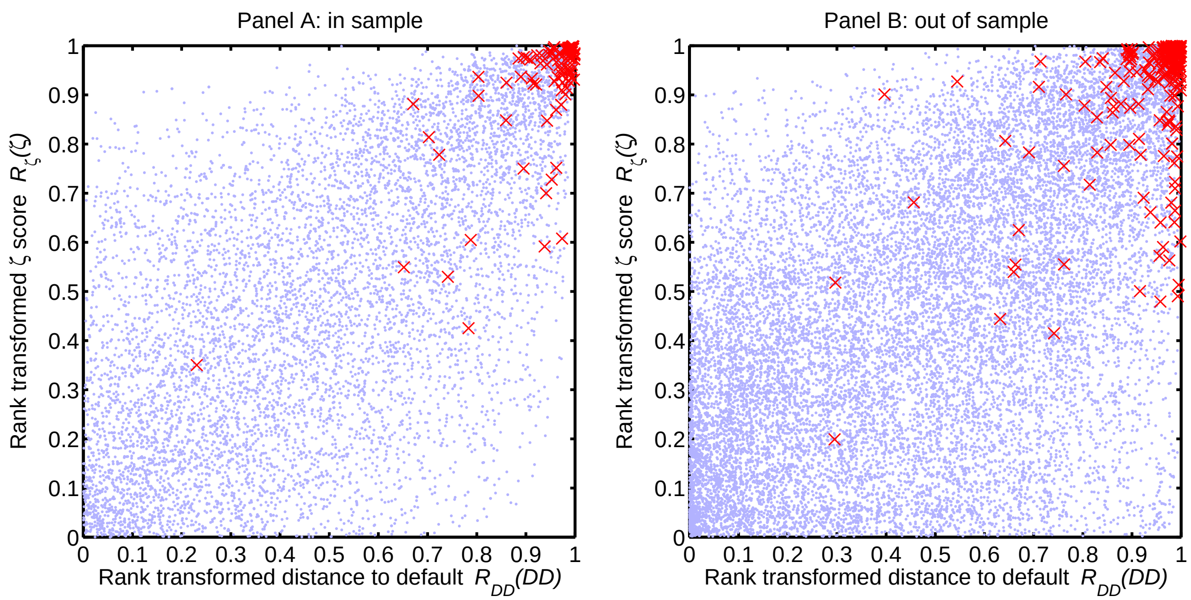

Figure 4. JOINT DISTRIBUTION OF ζ-SCORE AND DISTANCE TO DEFAULT DD
In Panel A, we plot the rank-transformed ζ-score and distance to default DD for the estimation sample. In Panel B, we plot the same graph for the validation sample. These two credit measures are correlated. Spearman's rank correlation shows a coefficient of 64.1% in the estimation sample and 57.1% in the validation sample. Most defaults are concentrated in the highest decile.

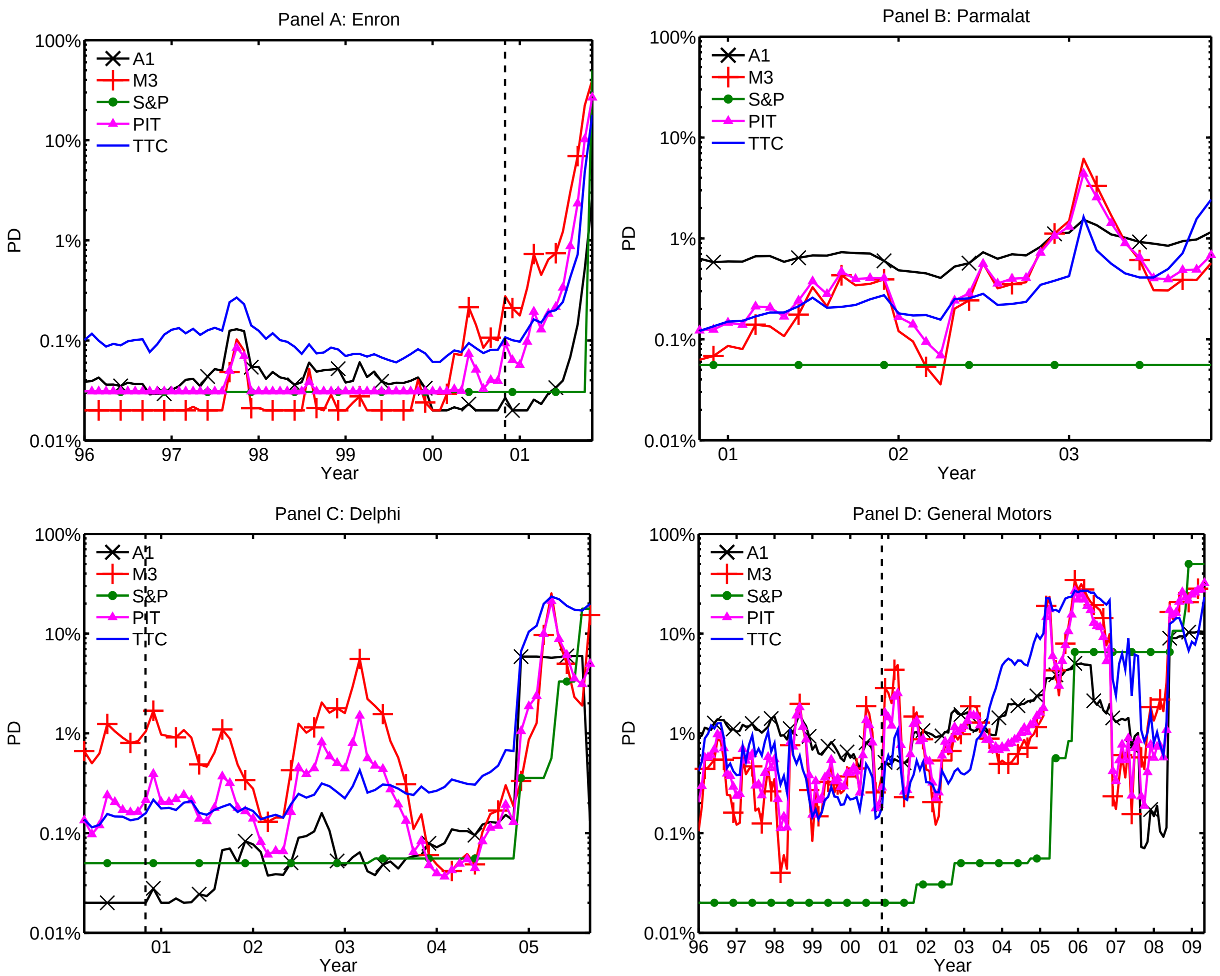

Figure 5. Defaulting Companies
One-year ahead default probability forecasts for the defaulting companies Enron, Parmalat, Delphi, and General Motors. The estimation sample period used for calibrating the models ranges from 1982 to 1999. The forecasts starting from 10/2000 are all out-of-sample. Enron went bankrupt in December 2001, Parmalat defaulted in December 2003, Delphi filed for Chapter 11 bankruptcy protection in October 2005, and the Chapter 11 filing of GM took place in June 2009.

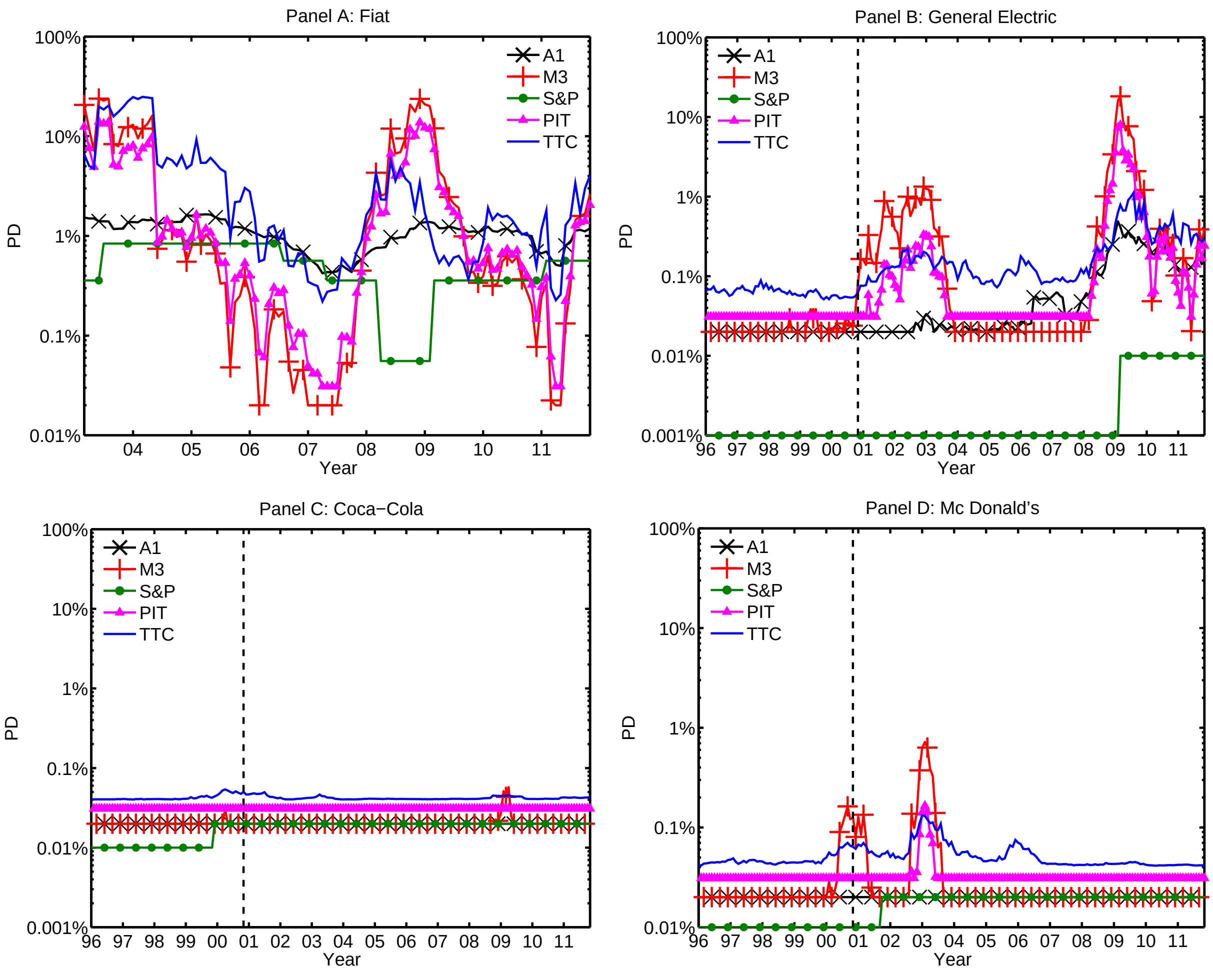

Figure 6. Non-Defaulted Companies
One-year-ahead default probability forecasts for non-defaulting companies: Fiat, General Electric, Coca-Cola, and McDonald's. The estimation sample period used for calibrating the models is 1982–1999. The forecasts starting from 10/2000 are all out-of-sample.

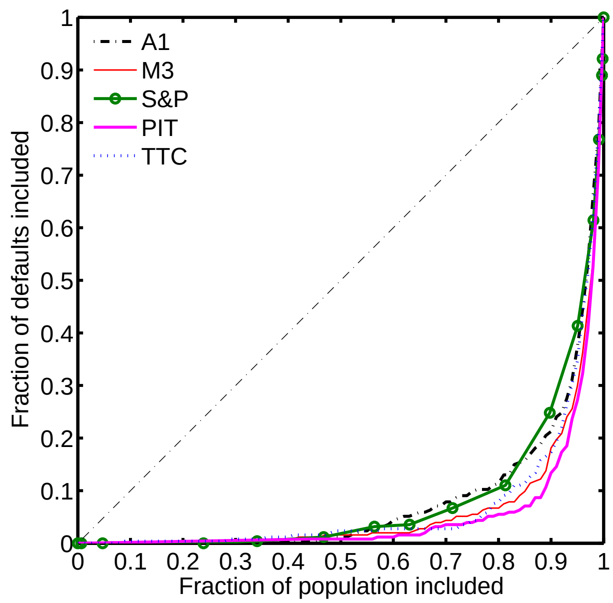

Figure 7. Lorenz Curve for Discrimination Testing
Empirical Lorenz curves for the various models for the validation sample from 10/2000 to 10/2010. Based on the empirical curves, **PIT** is more powerful than Altman's ζ-score, **A1**, Merton's distance to default, **M3**, and S&P's corporate issuer rating.

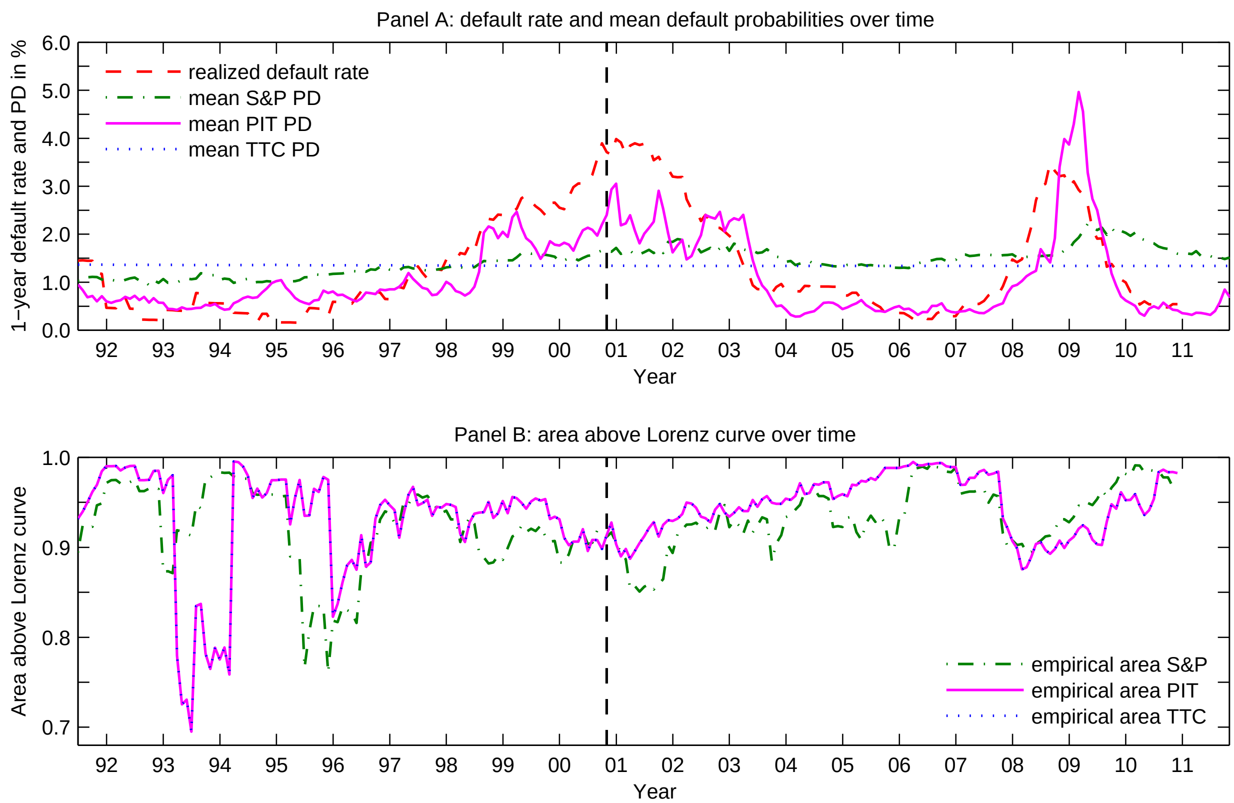

Figure 8. PREDICTED VERSUS REALIZED DEFAULT RATE AND DISCRIMINATORY POWER
Panel A compares the realized 1-year default rate to the mean PD predicted one year earlier. The dash-dotted line displays the mean PD computed from S&P's "through the cycle probabilities," the solid line displays the mean PD computed from model **PIT**, the dotted straight line is the mean PD of model **TTC** (which, by construction, is constant over time), the dashed line is the realized default rate. Panel B shows the empirical area over the Lorenz curve for S&P ratings, model **PIT**, and **TTC**, also on a monthly basis for default predictions over a 1-year forecast horizon. The vertical dashed line separates the estimation sample from the validation sample.

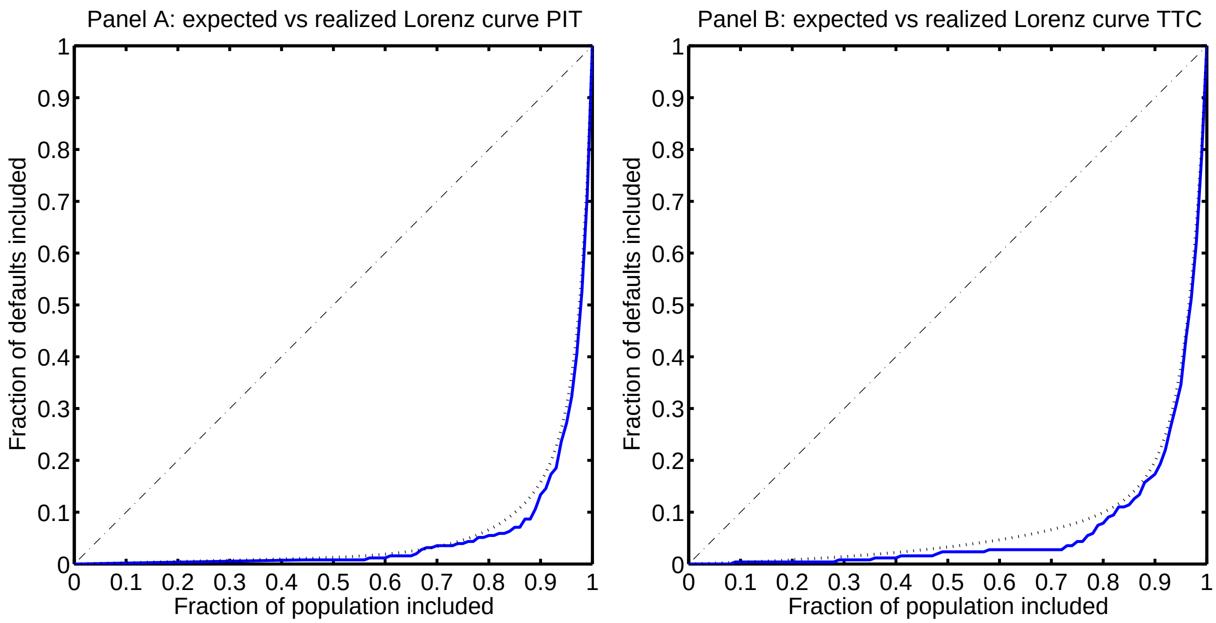

Figure 9. Lorenz Curves for Shape Calibration Testing
The empirical and the expected Lorenz curves of models **PIT** and **TTC**. Under the null hypotheses of shape calibrated forecasts, the empirical curves converge towards the expected curve with increasing sample size. Both models are shape calibrated.

Notes

[1]Some suggest simply replacing agency ratings with credit default swap (CDS) prices. In this line, Chava, Ganduri, and Ornthanalai (2012) find that markets react less to credit rating downgrades when the firm has CDS trading on its debt. However, CDS spreads incorporate a premium for systematic risk, which induces a bias in the default forecast. Two obligors with equal default probabilities can have different CDS spreads, depending on the economic state in which they tend to fail, see Coval, Jurek, and Stafford (2009), or Blöchlinger (2011).

[2]For instance: "Standard & Poor's credit ratings are not exact measures of the probability that a certain issuer or issue will default but are instead expressions of the relative credit risk of rated issuers and debt instruments." As of November 2012, Moody's writes on its web site: "Moody's ratings represent the opinion of Moody's Investors Service as to the *relative* creditworthiness of securities.". Similary, Fitch: "Credit ratings are opinions on *relative* credit quality and not a predictive measure of specific default probability."

[3]Hosmer and Lemeshow's goodness-of-fit test divides observations into deciles from lowest to highest risk based on estimated probabilities, then computes a χ^2-statistic from the observed and expected frequencies, and a p-value is computed from the χ^2-distribution to test the fit of the model.

[4]Duffie, Saita, and Wang (2007) assume an autoregressive Gaussian process of order one for the two-dimensional process of distance to default and log asset value, and they impose similar restrictions on the three-month and ten-year Treasury rate process as well as on the S&P return process.

[5]The joint notice issued by the Board of Governors of the Federal Reserve System, the Federal Deposit Insurance Corporation, and the Office of the Comptroller of the Currency, requests that the replacement of agency ratings in regulation requires that a potential alternative must be sufficiently transparent, unbiased, replicable, and defined so as to allow banking organizations of varying size and complexity to arrive at the same assessment of creditworthiness for similar exposures and to allow appropriate supervisory review. See the Federal Register Volume 76, Number 245 (Wednesday, December 21, 2011), `http://www.gpo.gov/fdsys/pkg/FR-2011-12-21/html/2011-32073.htm`.

[6]We will resort to the classical probit analysis as described by Maddala (1983) to obtain estimates

of (4). We also performed probit estimations including random effects to account for default dependencies. However, the resulting PD estimates do not change dramatically and our estimates without random effects perform just as well in the validation sample. Further, the academic literature, such as Shumway (2001), Chava and Jarrow (2004), and Campbell, Hilscher, and Szilagyi (2008), is also based on logistic regressions without random effects.

[7]We use the same database as in Blöchlinger (2012), which was extended by one year and enriched by the Z-score and ζ-score variables. We thank Basile Maire for augmenting the dataset.

[8]The influence of accounting data can vary over a quarter as in the model of Duffie and Lando (2001), newly released data seems slightly more relevant than stale accounting information towards the end of the quarter. However, our main results regarding discrimination and calibration remain unchanged if we choose another reference month. In particular, we performed the equivalent empirical analysis for end of November data.

[9]See, e.g., Altman and Rijken (2006) for a discussion of the difference between the point-in-time and through-the-cycle approaches.

[10]See the amendments to rule 15c3-1, `http://www.sec.gov/rules/proposed/2011/34-64352.pdf`.

[11]See, e.g., Association for Financial Professionals (2002)

[12]A prominent example of such an epidemic effect is the downgrade of AIG in 2008, leading to multiple collateral calls, increased liquidity stress, and falling market confidence.

[13]To derive test statistics on discrimination and calibration, we will consider default dependencies under the assumptions of Blöchlinger (2012). In particular, we assume that default dependence within one year is driven by a single factor following a β-distribution. Conditional on this factor, defaults are independent and Bernoulli distributed. The dependency is completely specified with two parameters, the factor loading ω, and the factor volatility σ. We choose $\omega = 0.8$, and $\sigma = 0.7889$. For details regarding the choice of these parameters, and technical details of the tests, we refer to Blöchlinger (2012).

[14]To clarify, we give a simple example: Suppose that a credit portfolio consists of two borrower groups of equal size, one group with a PD estimate of $1/2\,p$, the other group with a PD estimate of

$3/2\,p$ with $p \in (0, 2/3)$. In this case, we expect to observe 25% of all defaults in the first group and 75% in the second group, independently of the PD level p. If the true PD ratio between the two groups is indeed 3 to 1, then the expected Lorenz curve coincides with the true Lorenz curve.

[15]As a mnemonic device, we use $\mathbb{A} + \mathbb{B}$ to denote $\mathbb{A} \cup \mathbb{B}$ when $\mathbb{A}$ and $\mathbb{B}$ are disjoint.

[16]For each of the following currencies, we have interest rates available: AUD, BRL, CAD, CHF, EUR, GBP, HKD, IDR, INR, JPY, KRW, MXN, MYR, NOK, NZD, SEK, SGD, THB, USD. For the remaining currencies, we use the interest rates in parentheses: ARS (USD), CLP (USD), CZK (EUR), DKK (EUR), INR (USD), KWD (USD), TRY (EUR), TWD (JPY).

Table I
Data Summary

We divide our data into an estimation sample covering the period from 1982 to 1999, and a validation sample with data from 2000 to 2010. Our S&P ratings data end in October 2011, hence the last date for which we can analyze whether a given company defaults within one year is October 2010. The estimation sample contains 1,581 firms, the validation sample 2,549 firms. There are 347 firms that defaulted, 93 are in the estimation sample, 254 in the validation sample. We have 24,411 firm-year observations in total.

Industry	Total firms		Total firm-years		Defaults	
	1982–1999	2000–2010	1982–1999	2000–2010	1982–1999	2000–2010
Aerospace/Automotive/Capital Goods/Metal	258	425	1,297	2,982	9	51
Consumer/Service Sector	298	446	1,624	2,983	28	45
Energy and Natural Resources	139	245	614	1,511	8	12
Forest & Building Products/Homebuilders	105	137	553	952	9	21
Health Care/Chemicals	171	280	880	1,858	8	14
High Tech/Computers/Office Equipment	135	221	686	1,273	8	9
Leisure Time/Media	185	232	738	1,517	13	29
Real Estate	8	37	14	217	0	1
Telecommunications	78	166	264	860	4	45
Transportation	69	124	341	803	6	20
Utility	135	236	673	1,771	0	7
Total	1,581	2,549	7,684	16,727	93	254

Table II
Estimation of Merton's Model, Altman's Model and Combinations

Results from different probit regression models that transform the two ordinal credit risk measures distance to default and ζ-score into a 1-year default probability. The corresponding t-values are reported in brackets. We denote by '*' and '**' significance at the 5% and 1% level, respectively. The estimates are derived from our estimation sample covering the S&P corporate data base for the period from 1982 to October 1999.

Variable	**M1**	**M2**	**M3**	**M4**	**M5**	**A1**	**PIT**	**TTC**
Constant	0	-1.57**	-4.49**	-3.66**	-4.96**		-2.88**	-3.37**
		(-29.54)	(-10.79)	(-10.73)	(-10.99)		(-5.41)	(-13.25)
DD	-1	-0.34**						
		(-13.46)						
$R_{DD}(DD)$			2.64**	1.06*	3.33**		1.79**	
			(5.17)	(2.16)	(6.25)		(3.41)	
$R_{DD}(DD)^p$			1.63**	2.04**	1.56**		1.27**	
			(7.36)	(7.79)	(6.80)		(5.39)	
$r_{t,S}(S)$								0.83*
								(2.09)
$r_{t,S}(S)^p$								1.99**
								(8.08)
Z						-1.00**	-0.40**	
						(-38.42)	(-4.21)	
Exponent p		0	30	10	50		30	10
Log likelihood	-1158.02	-330.31	-321.32	-321.70	-324.65	-351.2	-311.71	-332.46

Table III
Estimation of Z-score and ζ-score

We estimate several probit regression models. To limit the influence of outliers, we either Winsorize the data at the 2.5% and 97.5% quantiles to estimate the Z-score and the ζ-score, or we rank transform the data, or we transform the data twice by first applying a rank transformation and then second a square-root transformation. The corresponding t-values are shown in brackets. We denote by '*' and '**' significance at the 5% and 1% level, respectively.

	Z-score (Wins.)	Z-score (rank tr.)	ζ-score (Wins.)	ζ-score (rank tr.)	ζ-score (rank tr.)	ζ-score (twice tr.)
Constant	-2.00**	-1.02**	-1.64**	-0.55**	-0.73**	0.22
	(-21.48)	(-8.05)	(-12.23)	(-3.50)	(-7.03)	(1.35)
Z_1: WorkCap/Assets	-0.49	-0.29				
	(-1.76)	(-1.68)				
Z_2, ζ_4: Profitability	-0.76**	-1.08**	-3.99**	-1.07**	-0.77**	-0.99**
	(-5.32)	(-4.43)	(-5.11)	(-3.68)	(-2.66)	(-3.45)
Z_3, ζ_1: EBIT/assets	-4.72**	-1.30**	0.23	0.32*		
	(-6.86)	(-5.30)	(1.67)	(2.03)		
Z_4: Leverage	-0.14**	-1.81**				
	(-2.98)	(-6.52)				
Z_5: Sales/assets	0.16**	0.45**				
	(2.54)	(2.72)				
ζ_2: Earn. stability			0.00	0.11		
			(1.16)	(-0.35)		
ζ_3: Debt service			-0.71**	-0.98**	-1.01**	-0.67**
			(-4.38)	(-3.70)	(-3.91)	(-2.62)
ζ_5: Current ratio			-0.12*	-0.48**		
			(-2.54)	(-2.91)		
ζ_6: Capitalization			-0.54**	-1.63**	-1.62**	-1.88**
			(-4.93)	(-6.20)	(-6.40)	(-8.21)
ζ_7: Relative log size			-0.19**	-1.41**	-1.28**	-1.31**
			(-5.60)	(-6.54)	(-6.12)	(-6.30)
Log likelihood	-407.63	-376.85	-367.60	-355.28	-368.74	-351.20

Table IV
Discriminatory Power

Discriminatory power of different forecasting models. The observations are ranked in descending order according to credit quality. The first column reports the fraction of the population included. For a given fraction of the population, we list the fraction of defaults. We list the corresponding percentages for the different models in the subsequent columns.

Fraction of	Fraction of defaulters included				
population included	**A1**	**M3**	S&P	**PIT**	**TTC**
50.00%	1.18%	1.57%	1.85%	0.79%	1.57%
75.00%	8.66%	5.51%	8.31%	3.94%	4.33%
80.00%	11.02%	6.69%	10.46%	5.12%	7.48%
85.00%	14.96%	9.84%	17.04%	7.09%	11.02%
90.00%	20.87%	18.09%	25.54%	12.60%	17.72%
95.00%	35.81%	29.84%	41.15%	26.76%	33.85%
96.00%	43.70%	36.22%	47.63%	31.89%	42.52%
97.00%	52.77%	44.89%	54.33%	40.56%	50.80%
98.00%	64.99%	53.91%	61.03%	52.42%	61.85%

Table V
Discrimination Testing: t-Values in Cross Tests

The table reports the discrimination tests for the out-of-sample statistical analysis. The first column reports the area above the Lorenz curve. The larger the area, the more powerful are the forecasts to separate defaulters from non-defaulters. The other columns report the t-values of testing the discrimination of different models against each other. We denote by '*' and '**' significance at the 5% and 1% level, respectively.

	Area above Lorenz curve	vs. naive	vs. **A1**	vs. **M3**	vs. S&P	vs. **PIT**	vs. **TTC**
A1	92.34%	59.130**	–	-2.194*	0.664	-4.998**	-1.559
M3	93.77%	65.437**	2.194*	–	2.255*	-4.962**	1.290
S&P	91.84%	58.854**	-0.664	-2.255*	–	-3.850**	-1.723
PIT	94.83%	78.075**	4.998**	4.962**	3.850**	–	4.448**
TTC	93.30%	63.857**	1.559	-1.290	1.723	-4.448**	–

Table VI

Calibration Testing of 1-Year Default Probabilities of Model PIT

Calibration tests for model **PIT**. We denote by '*' and '**' significance at the 5% and 1% level, respectively, with the corresponding p-values in brackets. The level statistic compares the realized default rate (real. PD) with the expected rate (exp. PD). The shape statistic depends on the difference between the realized (real. area) and the expected area (exp. area) above the Lorenz curve. Both statistics have a χ^2-distribution with one degree of freedom. The combined statistic is a summary statistic on level and shape (χ^2-distributed with two degrees of freedom). The p-values of the three statistics are shown in brackets. The last line shows the multi-period test statistic, the sum of the yearly statistics which follow a $\chi^2_{<11>}$- or $\chi^2_{<22>}$-distribution, respectively.

time period	obs.	def.	exp. PD	real. PD	exp. area	real. area	level	shape	comb.
10/2000–10/2001	1,213	45	2.40%	3.71%	0.916	0.917	0.940	0.002	0.942
							(0.332)	(0.961)	(0.624)
10/2001–10/2002	1,310	45	2.55%	3.44%	0.921	0.926	0.544	0.085	0.629
							(0.461)	(0.770)	(0.730)
10/2002–10/2003	1,360	29	2.48%	2.13%	0.919	0.951	0.001	2.556	2.557
							(0.976)	(0.110)	(0.278)
10/2003–10/2004	1,551	14	0.50%	0.90%	0.936	0.949	1.528	0.134	1.662
							(0.216)	(0.714)	(0.436)
10/2004–10/2005	1,653	15	0.57%	0.91%	0.946	0.954	1.027	0.055	1.081
							(0.311)	(0.815)	(0.582)
10/2005–10/2006	1,666	6	0.43%	0.36%	0.943	0.988	0.000	0.638	0.638
							(0.999)	(0.424)	(0.727)
10/2006–10/2007	1,724	6	0.42%	0.35%	0.942	0.990	0.000	0.678	0.679
							(0.984)	(0.410)	(0.712)
10/2007–10/2008	1,642	15	0.56%	0.91%	0.947	0.927	1.147	0.359	1.506
							(0.284)	(0.549)	(0.471)
10/2008–10/2009	1,619	52	3.40%	3.21%	0.893	0.909	0.030	0.801	0.831
							(0.862)	(0.371)	(0.660)
10/2009–10/2010	1,518	19	0.92%	1.25%	0.900	0.942	0.567	1.630	2.198
							(0.451)	(0.202)	(0.333)
10/2010–10/2011	1,471	8	0.43%	0.54%	0.924	0.984	0.432	1.325	1.757
							(0.511)	(0.250)	(0.415)
10/2000–10/2011	16,727	254	1.27%	1.52%	0.939	0.948	6.216	8.264	14.480
							(0.859)	(0.689)	(0.884)

Table VII

Calibration Testing: χ^2-Values for Level, Shape and Combined Statistics

The multi-period level, shape and combined statistics. We denote by '*' and '**' significance at the 5% and 1% level, respectively. All models except S&P provide calibrated forecasts, i.e., we cannot reject the combined multi-period calibration hypothesis for all but S&P. The underlying reason for S&P's miscalibration is that the PD ratio (=shape) between non-investment graded issuers and investment rated borrowers was significantly lower in 2002 than in 2010 (see Table VIII). The cycle adjusted version **TTC** provides calibrated forecasts even though that these forecasts are based only on relative risks and therefore disregard macroeconomic information.

	level $\chi^2_{<11>}$	shape $\chi^2_{<11>}$	combined $\chi^2_{<22>}$
A1	7.744	12.711	20.456
	(0.736)	(0.313)	(0.555)
M3	9.326	12.263	21.588
	(0.592)	(0.344)	(0.485)
S&P	19.141	30.571^{**}	49.711^{**}
	(0.059)	(0.001)	(0.001)
PIT	6.216	8.264	14.480
	(0.859)	(0.689)	(0.884)
TTC	21.365^{*}	9.256	30.621
	(0.030)	(0.598)	(0.104)

Table VIII

Calibration Testing of 1-Year Default Probabilities of S&P

Calibration tests for S&P ratings. We denote by '*' and '**' significance at the 5% and 1% level, respectively, with the corresponding p-values in brackets. The level statistic compares the realized default rate (real. PD) with the expected rate (exp. PD). The shape statistic depends on the difference between the realized (real. area) and the expected area (exp. area) above the Lorenz curve. Both statistics have a χ^2-distribution with one degree of freedom. The combined statistic is a summary statistic on level and shape (χ^2-distributed with two degrees of freedom). The p-values of the three statistics are shown in brackets. The last line shows the multi-period test statistic, the sum of the yearly statistics which follow a $\chi^2_{<11>}$- or $\chi^2_{<22>}$-distribution, respectively.

time period	obs.	def.	exp. PD	real. PD	exp. area	real. area	level	shape	comb.
10/2000–10/2001	1,213	45	1.65%	3.71%	0.913	0.912	2.731	0.007	2.738
							(0.098)	(0.933)	(0.254)
10/2001–10/2002	1,310	45	1.61%	3.44%	0.918	0.865	2.411	13.116**	15.527**
							(0.120)	(0.000)	(0.000)
10/2002–10/2003	1,360	29	1.71%	2.13%	0.924	0.939	0.378	0.681	1.059
							(0.539)	(0.409)	(0.589)
10/2003–10/2004	1,551	14	1.53%	0.90%	0.919	0.904	0.282	0.296	0.577
							(0.596)	(0.587)	(0.749)
10/2004–10/2005	1,653	15	1.33%	0.91%	0.912	0.923	0.101	0.148	0.250
							(0.751)	(0.700)	(0.883)
10/2005–10/2006	1,666	6	1.32%	0.36%	0.913	0.897	3.536	0.120	3.656
							(0.060)	(0.729)	(0.161)
10/2006–10/2007	1,724	6	1.44%	0.35%	0.910	0.984	4.945*	2.734	7.679*
							(0.026)	(0.098)	(0.022)
10/2007–10/2008	1,642	15	1.57%	0.91%	0.904	0.915	0.299	0.167	0.466
							(0.584)	(0.683)	(0.792)
10/2008–10/2009	1,619	52	1.58%	3.21%	0.906	0.931	2.129	2.844	4.973
							(0.145)	(0.092)	(0.083)
10/2009–10/2010	1,518	19	2.02%	1.25%	0.908	0.973	0.206	7.396**	7.602*
							(0.650)	(0.007)	(0.022)
10/2010–10/2011	1,471	8	1.64%	0.54%	0.901	0.967	2.123	3.061	5.184
							(0.145)	(0.080)	(0.075)
10/2000–10/2011	16,727	254	1.57%	1.52%	0.912	0.918	19.141	30.571**	49.711**
							(0.059)	(0.001)	(0.001)